

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
EVOLUTSIOONILISE BIOLOOGIA ÕPPETOOL

**Vanas DNAs leiduva saastuse tuvastamine vanade ja kaasaegsete mitogeenide
võrdlusanalüüsi abil**

Bakalaureusetöö

12 EAP

Maria Maloverjan

Juhendajad PhD Alena Kushniarevich

PhD Kristiina Tambets

TARTU 2018

INFOLEHT

Vanas DNAs leiduva saastuse tuvastamine vanade ja kaasaegsete mitogenoomide võrdlusanalüüsi abil

Vana DNA (*ancient* DNA, aDNA) on surnud organismide kudedest eraldatud geneetiline materjal. Vana DNA järjestustest saadakse vahetut informatsiooni paljude kunagi elanud liikide geneetilise mitmekesisuse kohta, eriti palju teaduslikke töid on seotud inimese ja tema eellaste uurimisega. Üks DNA omadusi on selle ebastabiilsus: aja jooksul suurenevad selle lagunemisaste ja keemilise modifitseerituse määr. Need protsessid raskendavad vana DNA analüüsi efektiivsust võrreldes kaasaegse DNAGA.

Arheoloogiliste proovide saastumine kaasaegse geneetilise materjaliga on olnud suureks probleemiks kogu vana DNA uurimisajaloo vältel. Käesoleva töö eesmärgiks on DNA eraldamine Tartu Ülikooli arheoloogide bioloogilisest materjalist, arheoloogide mitokondriaalse DNA (mtDNA) fülogeneetiline analüüs ning kaasaegsete ja vanade mtDNA järjestuste võrdlusanalüüs kompleksse saastuse tuvastamise protseduuri osana.

Märksõnad: vana DNA, saastus, mtDNA, geneetiline võrdlusanalüüs

CERCS kood: B220 Geneetika, tsütogeneetika

Detection of ancient DNA contamination through comparative analysis of modern and ancient mitogenomes

Ancient DNA (aDNA) is a genetic material extracted from tissues of dead organisms. aDNA sequences provide the direct information about the genetic diversity of various species that lived in the past, and particularly wide field of scientific research is connected to the human and his ancestors. One of the properties of DNA is its poor chemical stability: with time the rate of degradation and chemical modifications of the molecule increase. These processes affect the effectiveness of aDNA analysis making it difficult in comparison with modern DNA.

Contamination of ancient samples with modern genetic material has been a serious problem throughout the whole history of aDNA studies. The aim of the present work is to extract DNA from biological samples of archaeologists of University of Tartu, who have provided bone material for further aDNA analysis, to perform the phylogenetic analysis of archaeologists' mtDNA and carry out a comparative analysis of modern and ancient mtDNA sequences as part of a complex procedure of aDNA contamination detection.

Keywords: aDNA, contamination, mtDNA, comparative genetic analysis

CERCS code: B220 Genetics, cytogenetics

SISUKORD

INFOLEHT	2
SISUKORD	3
Kasutatud lühendid.....	5
Sissejuhatus	6
1 Kirjanduse ülevaade	7
1.1 Inimese mineviku uurimine geneetiliste meetodite abil.....	7
1.1.1 Genoomi haploidsed osad ja nende roll kaasaegsete inimpopulatsioonide evolutsiooni uurimisel	7
1.1.2 Autosoomid ja nende mitmekesisus kaasaegsetes inimpopulatsioonides	10
1.2 Vana DNA kui inimajaloo uurimise vahend.....	12
1.2.1 Inimese vana DNA allikad	13
1.2.2 DNA degradatsioon ja selle sõltuvus keskkonnatingimustest	15
1.2.3 Vana DNA eraldamine ja analüüsi spetsiifika	17
1.2.3.1 Vana DNA analüüsi varased meetodid	17
1.2.3.2 Vana DNA molekulaargeneetiline analüüs tänapäeval.....	18
1.3 Vana DNA saastus	20
1.3.1 Saastuse peamised allikad ja põhjused.....	21
1.3.2 Vana DNAGA töötamise peamised reeglid ja kriteeriumid	22
1.3.3 Saastuse tuvastamise viisid	23
1.3.3.1 Kontrollproovid.....	23
1.3.3.2 Sekveneerimisfragmentide pikkuse analüüs	24
1.3.3.3 Surmajärgsete modifikatsioonide analüüs.....	24
1.3.3.4 Vana DNA võrdlemine kaasaegse DNAGA	25
1.4 Saastuse tagajärjed	26
2 Eksperimentaalosa.....	28
2.1 Töö eesmärgid.....	28
2.2 Materjalid ja meetodika	28

2.2.1 Kasutatud proovid ja üldine töö käik	28
2.2.2 DNA eraldamine	28
2.2.2.1 DNA eraldamine põseepiteeli kaapmest	28
2.2.2.2 DNA eraldamine süljeproovidest.....	29
2.2.2.3 DNA kontsentratsiooni mõõtmine	30
2.2.2.4 Proovide kontsentreerimine	30
2.2.2.5 DNA täiendav puhastamine	30
2.2.3 Täisgenoomne mtDNA järjestuse määramine	31
2.2.3.1 Pikkade fragmentide amplifitseerimine	31
2.2.3.3 Amplifitseerimisprodukti visualiseerimine geelelektroforeesi abil	32
2.2.3.4 Amplifitseerimisprodukti puhastamine.....	32
2.2.3.5 Sekveneerimine	32
2.2.4 Lühikeste fragmentide amplifitseerimine ja sekveneerimine	33
2.2.5 Andmete analüüs.....	36
2.3 Tulemused ja arutelu	37
2.3.1 Arheoloogide mtDNA haplogrupid	37
2.3.2 Arheoloogide mtDNA võrdlus Eesti kaasaegse elanikkonna andmestikuga .	38
2.3.3 Arheoloogide mtDNA võrdlus Eesti keskaegse elanikkonna andmestikuga .	39
Kokkuvõte	40
Summary	41
Kasutatud kirjandus.....	42
Kasutatud veebiaadressid	49
Lisad.....	50
Lisa 1.....	50
Lisa 2.....	52
LIHTLITSENTS	53

Kasutatud lühendid

aDNA – *ancient DNA*, vana DNA

AMI – anatoomiliselt moodne inimene

ap – aluspaar

EDTA – etüleendiamiin-tetraädikhape

F ja R – *forward primer and reverse primer*, päripidine praimer ja äraspidine praimer

hg – haplogrupp ehk (antud töös) monofüleetiline mtDNA grupp (ehk klaad), mis hõlmab kõiki sarnaseid ühiselt eellaselt pärinevaid mtDNA tüüpe (haplotüüpe)

HVS – *hypervariable segment*, hüpervarieeruv ala

IBD – *identity by descent*, identse päritoluga (segmendid)

mtDNA – mitokondriaalne DNA

mt-MRCA – *mitochondrial most recent common ancestor*, mitokondriaalne viimane ühine eellane

Ne – *effective population size*, efektiivse populatsiooni suurus

NGS – *next generation sequencing*, järgmise põlvkonna sekveneerimine

np – nukleotiidne positsioon

NRY – *non-recombining region of Y-chromosome*, Y-kromosoomi mitterekombineeruv osa

PCA – *principal component analysis*, põhikomponentide analüüs

PCR – *polymerase chain reaction*, polümeraasi ahelreaktsioon

rpm – *revolution per minute*, pööret minutis

RSRS – *Reconstructed Sapiens Reference Sequence*, inimese mtDNA referentsjärjestus

UNG – uratsiil-DNA-glükosülaas

WISC – *whole-genome in-solution capture*, kogu genoomi “kinnipüüdmine” lahuses

Sissejuhatus

Vana DNA ehk aDNA (*ancient DNA*) on DNA, mida eraldatakse kümneid, sadu ja isegi sadu tuhandeid aastaid tagasi surnud organismide kudedest. Iidsetest inimjäänustest eraldatav DNA, nagu ka praegu elavate inimeste DNA, leiab laialdast kasutust evolutsioonilistes ja populatsioonilistes uuringutes. Üks suurimaid probleeme vana DNAGA seotud töö juures on saastus kaasaegse DNAGA, mille allikaks osutuvad kõige tihedamini arheoloogilise materjaliga töötavad inimesed.

Vaatamata sellele, et saastuse minimeerimiseks on välja töötatud rida reegleid ja ettevaatusabinõusid, ei ole selle täielik vältimine võimalik. Seetõttu on hädavajalik iidsete järjestuste autentsuse hoolikas kontrollimine. Autentsuse tõestamiseks kasutatakse nii vana DNA enda molekulaarseid eripärasid kui ka võrdlusanalüüsi: vana DNAd võrreldakse selle võimalikeks saasteallikateks olevate inimeste DNAGA.

Käesoleva töö eesmärgiks on vanade mitokondriaalse DNA (mtDNA) järjestuste autentsuse kontrollimine võrdlusanalüüsi abil, mis on osa komplekssest vana DNA saastuse tuvastamise protseduurist. Vana DNA, mille autentsust kontrollitakse, pärineb eesti arheoloogide rühma poolt läbiviidud väljakaevamiste käigus saadud luumaterjalist. Töö eesmärgi saavutamiseks olid püstitatud järgmised ülesanded: eraldada arheoloogide DNA kahest erinevast bioloogilisest materjalist (süljest ja põseepiteeli kaapmest); viia läbi DNA proovide kvalitatiivne ja kvantitatiivne hinnang; määrata mtDNA haplogrupid kogu mitokondri genoomi või mtDNA kontrollregiooni sekveneerimise ja kodeeriva regiooni informatiivsete positsioonide genotüpeerimise alusel; kõrvutada arheoloogide mtDNA haplogruppe üldise mtDNA mitmekesisusega eestlaste tänapäevases populatsioonis ning teostada vana DNA ja arheoloogide DNA võrdlusanalüüs.

Töö sisaldab ka teoreetilist osa, mis kujutab endast vana DNAd ja mtDNAd käsitleva kirjanduse ülevaadet. Ülevaate eesmärgiks on anda ettekujutus vana DNA kasutamisest teaduslikes uuringutes, selle analüüsi eripäradest, saastuse problemaatikast ning sellega võitlemise võimalustest.

1 Kirjanduse ülevaade

1.1 Inimese mineviku uurimine geneetiliste meetodite abil

Populatsiooni- ja evolutsioonigeneetika alased uuringud võimaldavad rekonstrueerida nii iidseid protsesse (inimliigi teket, divergentsi ja segunemist teiste hominiididega, erinevate maailmaosade asustamist inimese poolt) kui ka hilisemaid mikroevolutsioonilisi sündmusi (migratsioone, arvukuse kõikumisi, populatsioonidevahelisi interaktsioone, keskkonnatingimustega kohastumist). Inimese evolutsioonilise mineviku uurimiseks kasutatakse nii praegu elavate kui ka ammu surnud indiviidide geneetilist materjali. Varieerub ka analüüsitava geneetilise materjali liik: selleks võib olla nii diploidne autosomaalne DNA, kui ka haploidsed, s.t. rakus ainsa variandina esindatud Y-kromosoom ja mtDNA. Nii autosoomidel, kui ka genoomi haploidsetel osadel on olemas eripärad, mis määravad ühe või teise eelistatust erinevate uuringute jaoks.

1.1.1 Genoomi haploidsed osad ja nende roll kaasaegete inimpopulatsioonide evolutsiooni uurimisel

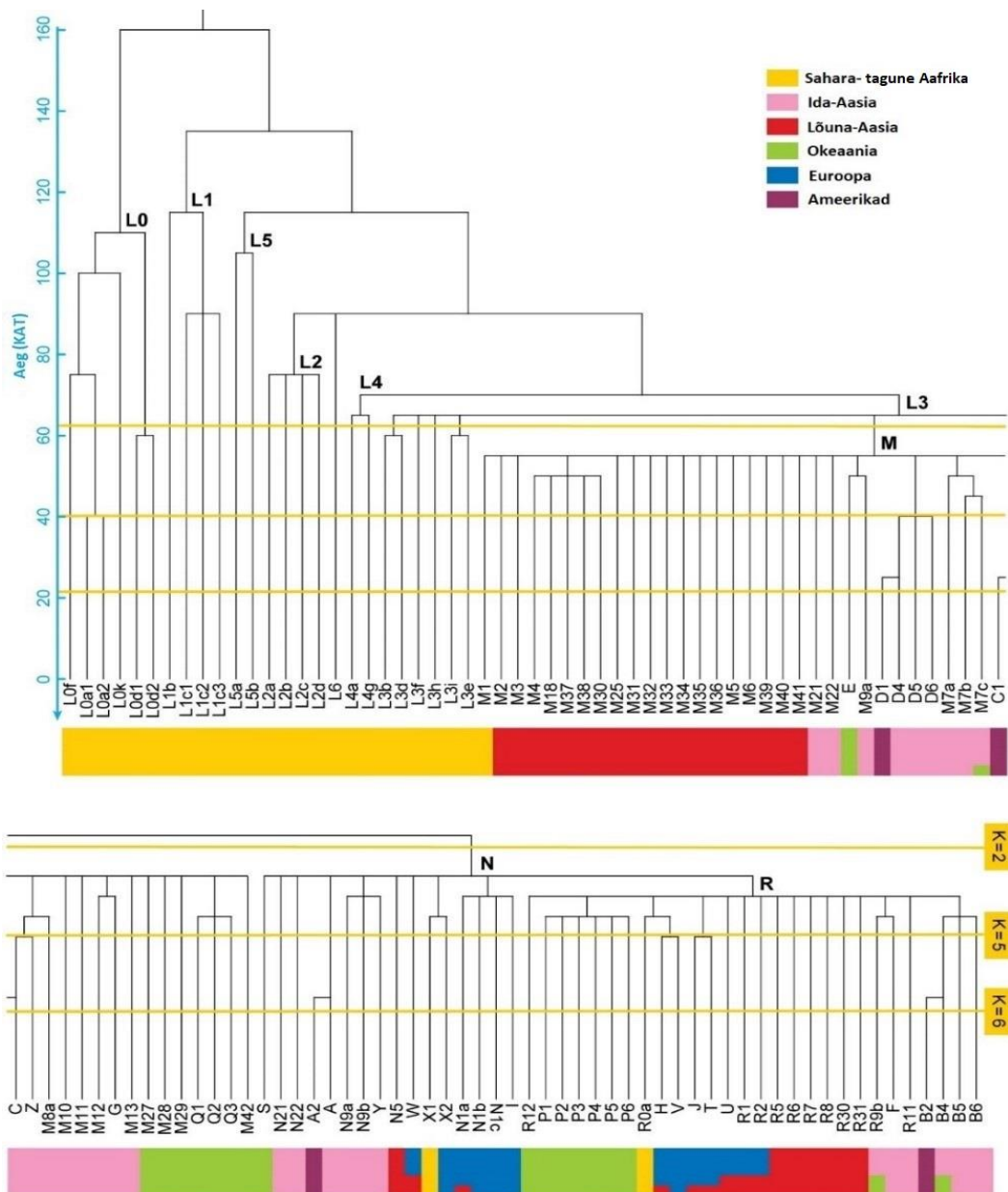
MtDNA ja mitte-rekombineeruv Y-kromosoomi osa (*non-recombining region of Y chromosome*, NRY) mängivad olulist rolli inimese mineviku uurimises tänu omadustele, mis eristavad neid autosoomidest. Mõlemad päranduvad uniparentaalselt (mtDNA ema- ja Y isaliini pidi) ega rekombineeru idutee rakkudes (Giles jt., 1980). Tänu rekombinatsiooni puudumisele pärandub mtDNA muutumatuna emalt lastele ning NRY isalt poegadele, mis teeb neid populatsioonide nais- ja meesliini rekonstrueerimise unikaalseteks vahenditeks. MtDNA ja NRY pärandumisviis võimaldab nende järjestused ühendada vastavalt ühte süsteemi ehk konstrueerida mtDNA ja Y-kromosoomi fülogeneetilised puud, mis kajastavad monofüleetiliste gruppide ehk haplogruppide vahelisi seoseid ja võimaldavad nende vanuse määramist. Veel üheks mtDNA ja Y-kromosoomi eripäraks võrreldes ülejäänud genoomiga on nende suurem evolutsioonikiirus (Brown jt., 1979), seetõttu kajastuvad nendes hästi ka evolutsiooniliselt suhteliselt lühikese aja jooksul toimunud muutused. Genoomi haploidsete osade kiirema evolutsiooni põhjusteks on kõrge mutatsioonisagedus ja neli korda väiksem efektiivse populatsiooni suurus (N_e) võrreldes autosoomidega (Jorde jt., 1998). Neljakordne vahe tuleneb sellest, et mtDNA ja Y-kromosoomi ülekandmisesse järglastele on kaasatud vaid pool populatsioonist (vastavalt naised ja mehed) ning ka nende rakkudes on mtDNA ja Y-kromosoom haploidsetel kujul, s.t. kaks korda väiksema variatsioonide arvuga, kui diploidsetel autosoomidel.

Inimese mtDNA, mida on lähemalt käsitletud käesoleva töö eksperimentaalosas, on kaheaheelaline tsirkulaarne molekul pikkusega umbes 16 569 aluspaari (ap) (Anderson jt., 1981, Andrews jt., 1999). See paikneb väljaspool rakutuuma mitokondrites ja replitseerub tuumagenoomist sõltumatult. MtDNAs eristatakse kaht regiooni: kodeerivat regiooni ja nn. kontrollregiooni. Kodeerivas regioonis asuvad geenid, millest ekspresseeritakse tRNAsid, rRNAsid, ribosoomi valke, oksüdatiivses fosforüleerimises osalevaid valke ja muid molekule. Mittekodeeriv kontrollregioon (nukleotiididsed positsioonid (np) 16 024 – 576) hõlmab umbes 7% kogu mtDNA pikkusest ning sisaldab kahte nn. hüpervarieeruvat ala (*hypervariable segment*, HVS), mida nimetatakse nii nende kõrge varieeruvuse tõttu inimpopulatsioonide vahel ja sees: HVSII (np 57 – 372) ja HVSI (np 16 024 – 16 383) (Andrews jt., 1999, Anderson jt., 1981, Giles jt., 1980, [1]).

MtDNA kõrget mutatsioonikiirust soodustavad lokaalselt madal pH mitokondrites, sage replikatsioon ja madal reparatsiooni efektiivsus (Birky, 2001). See kiirus ei ole konstantne kogu mtDNA ulatuses varieerudes sõltuvalt alast: kontrollregioon evolutsioneerub kiiremini, kui kodeeriv, ning ka HVS-piirkondade sees esineb tugev kohast sõltuv mutatsioonikiiruse varieeruvus (Pesole jt., 1999). Veel üheks mtDNA eripäraks on selle kõrge koopiaarv rakus. Igas mitokondris on kuni 15 DNA molekuli ja rakus, sõltuvalt selle energiavajadusest, võib kokku olla 1000 kuni 10 000 mtDNA koopiat (Rooney jt., 2015). Koos mtDNA suhteliselt väikese suurusega lihtsustab suur koopiate hulk olulisel määral selle molekulaargeneetilist analüüsi. Peab märkima ka seda, et vaatamata uniparentaalsele pärandumisele esineb mtDNA, erinevalt Y-kromosoomist, iga inimese rakkudes, võimaldades uurida kõiki populatsiooniliikmeid.

Eelpool loetletud mtDNA omadused muudavad selle populatsioonigeneetiliste uuringute hinnatud vahendiks. MtDNA on nii kaasaegsete (nt. Bandelt jt., 1995, Brandstätter jt., 2006), kui ka iidsete (nt. Krings jt., 1997, Adachi jt., 2017, Currat ja Excoffier, 2004) inimpopulatsioonide uurimises laialdast kasutust leidnud. Tasub mainida, et kõige esimesed surnud organismidele kuuluva DNA geneetilise analüüsi katsed põhinesid just mtDNA kontrollregioonil (Higuchi jt., 1984) ning aastal 2000 ilmunud töö, kus oli edukalt määratud 53 inimese mtDNA täielik järjestus, tähistas üleminekut kontrollregiooni sekveneerimisest kogu mtDNA järjestuse kasutamiseks evolutsioonilistes uuringutes (Ingman jt., 2000).

Tänapäevaks on kogutud suur hulk informatsiooni mtDNA ja Y-kromosoomi mitmekesisuse kohta kaasaegsetes inimpopulatsioonides. Mõlema jaoks on konstrueeritud fülogeneetilised puud (van Oven ja Klayser, 2009) [2] ning pidevalt toimub uute järjestuste sekveneerimine, millega kaasneb ka fülogeneetiliste puude täiustamine.



Joonis 1. MtDNA globaalne fülogeenia. Kõik Aafrika-välised mtDNA liinid lähtuvad vaid ühest haplogrupist (L3), mille mõned liinid on levinud ka Aafrikas. Kõik Aafrika-välised liinid kuuluvad vaid kolme suurde harusse (M, N ja selle haru R) (kohandatud Underhill ja Kivisild, 2007 järgi).

Fülogeneetiline ja fülogeograafiline mtDNA ja Y-kromosoomi analüüs annavad informatsiooni inimese ema- ja isapoolse geneetilise ajaloo kohta ja võimaldavad nende võrdlevat analüüsi. Näiteks esinevad mõlema lookuse vanimad põhiharud ainult Aafrikas ning vaid väikest osa kõikidest harudest võib kohata ülejäänud maailmaosades (joonis 1), mis toetab hilise Aafrikast väljarände teooriat (*recent „Out of Africa“ theory*), vastavalt millele oli Aafrika anatoomiliselt moodsa inimese (AMI) esmaseks levimiskoldeks enne teiste kontinentide asustamist. Lisaks on tänapäevaste mtDNA ja Y-kromosoomi järjestuste analüüsi alusel rekonstrueeritud hulk suhteliselt hiljutisi sündmusi inimese ajaloos, näiteks on modelleeritud

migratsiooniteid ja populatsioonidevahelisi interaktsioone peale Aafrikast väljumist. Samuti uuritakse haploidsete genoomi osade populatsioonidesisest ja -vahelist geneetilist mitmekesisust ning tehakse selle põhjal järeldusi mineviku demograafiliste trendide kohta: näiteks on näidatud, et geograafiliselt perifeersetes regioonides, nagu Euroopa ja Austraalia, oli väiksem hulk asutajaid, kui Aasias (Underhill ja Kivisild, 2007).

1.1.2 Autosoomid ja nende mitmekesisus kaasaegsetes inimpopulatsioonides

Inimese genoomi suuruseks on umbes 3.2 Gap (Brown, 2002), mis on kordades suurem, kui mtDNA (16,5 kap) ja NRY (57 Map) (Quintana-Murci ja Fellous, 2001). Seetõttu on autosoomid palju suurema hulga informatsiooni allikaks võrreldes haploidsete genoomi osadega. MtDNA ja Y-kromosoomi pikka aega püsinud madala N_e tulemusel koalestseeruvad mõlemad fülogeneetilised puud (s.t. kõik liinid sulanduvad üheks esivanemaks) suhteliselt hilises ajaperioodis võrreldes diploidse genoomiga. Inimliigi korral toimub koalestseerumine umbes 100 000 – 200 000 aastat tagasi, see piirab mtDNA ja Y-kromosoomi abil uuritavate protsesside jälgimist. Seega sobivad need eelkõige inimese ajaloo evolutsiooniliselt hiljutiste sündmuste uurimiseks. See-eest kogu genoomi uuringud heidavad valgust ka iidsematele protsessidele, mis on seotud inimliigi tekke ja mitmekesistumisega (Underhill ja Kivisild, 2007).

Erinevalt mtDNAst ja Y-kromosoomist ei kandu autosoomid muutumatuna vanemalt lastele. Tänu biparentaalsele pärandumisele, millega kaasneb homoloogiliste kromosoomide juhuslik lahknemine meioosi anafaasis ja rekombinatsioon ehk ristisiire meioosi esimeses profaasis, toimub pidev alleelide “segunemine” homoloogiliste kromosoomide vahel [3]. Seetõttu ei ole diploidseid autosomaalseid järjestusi, erinevalt mtDNA ja Y-kromosoomi järjestustest, võimalik jaotada terviklikeks haplogruppideks ja konstrueerida nende alusel ühtne fülogeneetiline puu. See-eest rakendatakse autosoomide uurimisel meetodeid, mis põhinevad kas ühenukleotiidsete asenduste sagedustel või haplotüüpide analüüsil. Näiteks kasutatakse esimesel juhul populatsioonide geneetilise struktuuri määramiseks F-statistikat, mille abil selgitatakse, kuivõrd erinevad on keskmised väärtused kahe populatsiooni jaoks (Wright, 1951, Weir ja Cockerham, 1984). Autosoomide abil saadava suure hulga andmestiku töötlemiseks kasutatakse ka põhikomponentide analüüsi (*principal components analysis*, PCA), mille peamiseks eesmärgiks on andmete esitusviisi lihtsustamine võimalikult väikeste informatsiooni kadudega (Pearson, 1901, Hotelling, 1933). Individuaalsete segunemisproportsioonide arvutamise üheks levinumaks meetodiks on segunemisanalüüs (*admixture analysis*, *biogeographical ancestry analysis*), mille abil määratakse indiviidi geograafiline päritolu tema

geneetilise tausta baasil. Meetod lähtub asjaolust, et indiviidi DNA kannab endas kõikide tema eellaspopulatsioonide geneetilist materjali (Alexander jt., 2009).

Vaatamata sellele, et sugurakkudes toimuv ristsiire põhjustab kromosoomide muutumist põlvkondade lõikes, toimub teatud ulatuslike regioonide vahel rekombinatsioon väga harva, nii et nad päranduvad reeglina muutumatu tervikuna, mistõttu saab analüüsida piirkondi, mis hõlmavad mitmeid varieeruvaid positsioone. Niisuguseid järjestusi nimetatakse haplotüüpideks (Crawford ja Nickerson, 2005) ning analüüsides neid erinevate populatsioonide esindajatel võib modelleerida ja dateerida mineviku segunemisprotsesse ja migratsioone (Hellenthal jt., 2014). Samal eesmärgil rakendatakse identse päritoluga segmentidel (*identity by descent*, IBD) põhinevat meetodit, mis on laialdaselt kasutatav geneetilises kaardistamises ja populatsioonigeneetilistes uuringutes: kahe indiviidi ühise IBD segmentide pikkus on seda suurem, mida hiljem elas nende viimane ühine eellane (Browning, 2008).

Praegu elavate inimeste autosoomide uurimine võimaldab teha järeldusi nii kaasaegsete populatsioonide struktuuri kui ka mineviku sündmuste kohta. Üheks heaks näiteks on 2008. aasta töö, kus uuriti geneetilise ja geograafilise distantssi korrelatsiooni ning genotüpeerides 3000 eurooplast üle 500 000 lookuse osas jõuti järeldusele, et kasutades alleelide sagedusi võib määrata indiviidi geograafilist päritolu mõnesaja kilomeetri täpsusega (Novembre jt., 2008). Peale selle teeb genoomne analüüs võimalikuks ajaloosündmuste rekonstrueerimise: näiteks modelleerisid Hellenthal ja kolleegid kaasaegsete inimeste autosomaalse genoomi baasil üle saja olulise sündmuse inimajaloos, mis toimusid viimase 4000 aasta vältel Lääne-Euraasia territooriumil (Hellenthal jt., 2014) ning mille seas on sõjad, sissetungid, migratsioonid, orjakauplemine ja muud populatsioonide segunemisega seotud protsessid. Veel üks kaasaegse DNA ulatuslik uuring viidi läbi 2016. aastal (Pagani jt., 2016). Sekveneerides täielikult ligi 500 erineva päritoluga inimese DNAd uurisid autorid heterosügootsust, geenitriivi, populatsioonide Ne muutumist ajas, positiivset ja tasakaalustavat looduslikku valikut ja ka inimese kaugemat minevikku: jõuti järeldusele, et lisaks AMI põhilisele väljarände Aafrikast 75 000 aastat tagasi leidis aset ka varasem väljaränne, mis toimus umbes 45 000 aastat enne seda.

Vaatamata kõikidele analüüsivõimalustele annab kaasaegsete inimeste DNA minevikusündmuste kohta vaid sellist informatsiooni, mis võimaldab küll püstitada hüpoteese meie liigi kunagise geneetilise mitmekesisuse kohta, aga ei anna võimalust nende kontrollimiseks. Seetõttu hakatakse populatsioonigeneetilistes ja evolutsioonilistes uuringutes aina rohkem kasutama ammu surnud inimeste geneetilist materjali. Sadu ja tuhandeid aastaid vanad DNA järjestused osutuvad tihtipeale asendamatuks vahendiks inimese mineviku mõistmisel.

1.2 Vana DNA kui inimajaloo uurimise vahend

Vanaks DNAs (*ancient* DNA, aDNA) nimetatakse DNAd, mis on eraldatud ammu surnud organismide kudedest. Vana DNA uurimisega tegeleb arheogenoomika, mis uurib surnud, sealhulgas väljasurnud organismide iseärasusi kasutades nende DNA järjestusi. Vana DNA abil uuritakse pajasid mineviku aspekte, näiteks kunagi elanud populatsioonide geneetilist mitmekesisust ja interaktsioone, pikaajalisi demograafilisi mustreid, keskkonnategurite (näiteks kliimamuutuste või katastroofide) ja evolutsiooni vahelisi seoseid.

Vana DNA uurimise alguspunktiks loetakse aastat 1984, kui Higuchi ja kolleegid sekveneerisid vana mtDNA lühikese fragmendi. Järjestus kuulus quaggale (sebrale sarnane imetaja, kes suri välja 1880ndatel aastatel) ning oli eraldatud muuseumikollektsiooni kuulunud nahast (Higuchi jt., 1984). Järgnes suur hulk vana DNA uuringuid, kus üritati eraldada ja sekveneerida DNAd, mis kuulus kõikvõimalikele väljasurnud organismidele (isegi dinosaurustele), sealhulgas ka hominiididele, kuid esialgu olid need katsed enamasti edutud või andsid ekslikke tulemusi.

Esimesed õnnestunud vana DNA uuringud põhinesid mtDNA1, mida on selle väikese suuruse ja suure koopiaarvu tõttu lihtsam eraldada ja analüüsida, kui tuuma DNAd (vt. punkti 1.1.1). Iidsetest jäänustest eraldatud mtDNA genoomi kattuvus (ehk mitu korda on iga nukleotiid DNA sekveneerimisel loetud) võib olla kümneid ja sadu kordi suurem, kui tuuma DNA oma (Hansen jt., 2017).

Kuid vana DNA eraldamise optimeerimise ja uue põlvkonna sekveneerimismeetodite (*next generation sequencing*, NGS) kasutuselevõtuga ei ole see vahe enam ületamatuks takistuseks nukleaarsete järjestuste saamisel. Tänu tehnoloogiate arengule toimus üleminek üksikute organismide mtDNA sekveneerimisest (nt. Higuchi jt., 1984, Pääbo jt., 1988) ulatuslikeks täisgenoomsetel järjestustel põhinevateks populatsiooniuringuteks, kus analüüsitakse kümneid ja sadu teatud kultuuri või regiooniga seotud genoomi (nt. Allentoft jt., 2015, Haak jt., 2015; Mathieson jt., 2016). Samuti sai võimalikuks erinevatest ajastutest pärinevate, isegi sadu tuhandeid aastaid vanade genoomsete järjestuste uurimine: näiteks sekveneeriti 2016. aastal 430 000-aasta vanusest hominiidi fossiilist eraldatud nukleaarne DNA (Meyer jt., 2016). Tänapäevaks on sekveneeritud suur hulk vanu genoomi, alates mikroobsetest patogeenidest ja lõpetades selgroogsetega (Orlando jt., 2015).

Vana DNA põhjal on teostatud rida inimese kaugminevikuga seotud revolutsioonilisi uuringuid. Näiteks sekveneeriti 1997. aastal nendertaallase mtDNA ning jõuti järeldusele, et AMI ja neandertaallase viimane ühine esivanem elas 4 korda varem kui nüüdisaegsete inimeste viimane ühine esivanem, mis viitab sellele, et neandertaallased surid välja ilma kaasaegse

inimese mtDNA mitmekesisusse panustamata (Krings jt., 1997). On ka sekveneeritud Denisova koopast leitud sõrmeluust ja hambast eraldatud DNAd (Reich jt., 2010, Meyer jt., 2012) ja rekonstrueeritud selle põhjal Denisova inimese demograafilist ajalugu ning täisgenoomse analüüsi põhjal jõutud järeldusele, et AMI, neandertaallane ja Denisova inimene on siiski kunagi segunenud (Green jt., 2010, Prüfer jt., 2014).

Tänapäevaks on kogunenud suur hulk informatsiooni inimese, eriti eurooplaste, iidsete populatsioonide kohta (nt. Mathieson jt., 2018) [4][5]. Selle informatsiooni analüüsimisel tehakse järeldusi nii inimese geneetilise mitmekesisuse kohta mingil kindlal ajastul kui ka DNA muutumisest põlvkondade ja ajastute lõikes.

Näiteks selgitati vana DNA põhjal tänapäevaste eurooplaste päritolu: on näidatud, et nende eellkäijaks on vähemalt kolm kunagist eellaspopulatsiooni: mesoliitikumi ehk keskmise kiviaja kütid-korilased, varased maaharijad ja stepipäritolu karjakasvatajad (Lazaridis jt., 2014; Haak jt., 2015). Vanade genoomide uurimine aitab kaasa Euroopa neolitiseerumise, s.t. hilisesse kiviaega üleminekul toimunud kultuuriliste ja rahvastiku muutustega seotud sündmuste mõistmisele (Bramanti jt., 2009, Skoglund jt., 2012, Mathieson jt., 2016): näiteks leidis kinnitust neoliitikumi karjakasvatajate ühise Anatoolia päritolu hüpoteesi enne nende edasist levimist Kesk-Euroopas.

Veel üheks uurimissuunaks on selliste DNA muutuste uurimine, mis kajastuvad fenotüübis ja on seeläbi loodusliku valiku all. Niisugusteks tunnusteks on näiteks naha ja silmade värvus, patogeenniresistentsus, immuunsüsteemi eripärad jne. Näiteks 2016. aasta uuringus genoomse vana DNA polümorfisme analüüsides näidati, et uuritud tunnustest on eurooplastel loodusliku valiku tugevaima mõju all laktoositaluvuse eest vastutav lookus ning et laktoositaluvus on eurooplastel välja kujunenud vaid viimase 4000 aasta jooksul. Samuti on valiku all teised toitainete (lipiidide, gluteeni ja vitamiin D) omastamisega seotud geenid ja ka naha heleda pigmentatsiooni eest vastutav lookus (Mathieson jt., 2016).

Niisiis täiendab inimese ja tema eellaste vana DNA uurimine olulisel määral kaasaegsele DNAlle toetuvat arusaamist minevikusündmustest. Analüüsides arheoloogilisest materjalist eraldatud DNA järjestusi saadakse vahetut informatsiooni ammu surnud indiviidide geneetilisest eripäradest ning tehakse järeldusi, mis puudutavad nii inimliigi teket ja seoseid teiste liikidega kui ka tänapäevaste populatsioonide suhteliselt hiljutist tekkelugu ja fenotüübiliste iseärasuste väljakujunemist.

1.2.1 Inimese vana DNA allikad

Vana DNA põhiliseks allikaks on inimjäänused, mida hoitakse muuseumikollektsioonides ja kirikutes või kaevatakse välja arheoloogiliste leidudena. Muuseumides hoitakse tavaliselt

skelette või muumiaid, kuid materjal võib olla esindatud ka mingil muul kujul – näiteks võivad kehad ja kehaosad olla fikseeritud formaliinis või piirituses (Thomas ja Pääbo, 1993). Suur hulk antropoloogilist materjali saadakse muistsete haudade lahtikaevamisel, kus jäänused on säilinud peamiselt luulise materjalina.

DNA eraldamiseks iidsete inimpopulatsioonide geneetilise analüüsi jaoks kasutatakse eelkõige just kaltsifitseerunud kudesid – luid ja hambaid. Parimaks DNA allikaks peetakse kaljuluud (*petrous bone* – oimuluu sisemine osa, mis on imetaja kehas kõige kõvem ja tahkem) ja hambajuure tsementi (Pinhasi jt., 2015, Hansen jt., 2017). Uuringud on näidanud, et mõlemast materjalist on võimalik eraldada keskmiselt 8 korda rohkem endogeenset, s.t. uuritavale organismile kuuluvat DNAd, kui teistest luudest (Hansen jt., 2017), mis tähendab, et DNA parema säilimise tõttu õnnestub samast hulgast koest saada järjestused keskmiselt 8x suurema kattuvusega. Kui võrrelda neid kahte, siis kaljuluud peetakse stabiilsemaks ja usaldusväärsemaks DNA allikaks, kuigi soodsate säilimistingimuste korral võib hammastest eraldada rohkem DNAd (Mathieson jt., 2016). Samas ei sobi kaljuluu üldiselt patogeenide uurimiseks – on näidatud, et näiteks katku põhjustav *Y. pestis*, mis on tuvastatav katku kandva inimese hambast, ei ole leitav kaljuluu materjalist (Margaryan jt., 2018).

Kuigi vana DNA eraldamiseks kasutatakse peamiselt väljakaevatud luulist materjali, kasutatakse ka küüsi, juukseid ja muid kudesid ja organeid, aga ka koproliite ehk kivistunud ekskrementide (Poinar jt., 1998). Näiteks eraldasid 2011. aastal Rasmussen ja kaasautorid DNAd Austraalia pärismaalase juustest (Rasmussen jt., 2011) ja ühes varasemas uuringus oli DNA allikaks 7000-aastane inimese aju (Pääbo jt., 1988). Veel üheks juuste kasutamise näiteks on varase ja tänaseks kadunud Gröönimaa kultuuri kuulunud ja 4000 aastat igikeltsas veetnud mehe juustest eraldatud mitokondriaalse (Gilbert jt., 2008) ja tuuma (Rasmussen jt., 2010) DNA sekveneerimine. Juuksed, nagu hambad ja kaljuluu, on samuti väga heaks DNA allikaks, kuna nende hüdrofoobne pind teeb neid vähem vastuvõtlikeks bakteriaalse kahjustuse suhtes ja võimaldab pinna puhastamist enne DNA eraldamist (Gilbert jt., 2004), kuid juustega jäänused on liiga haruldased selleks, et need võiks olla laialdaselt kasutatavad.

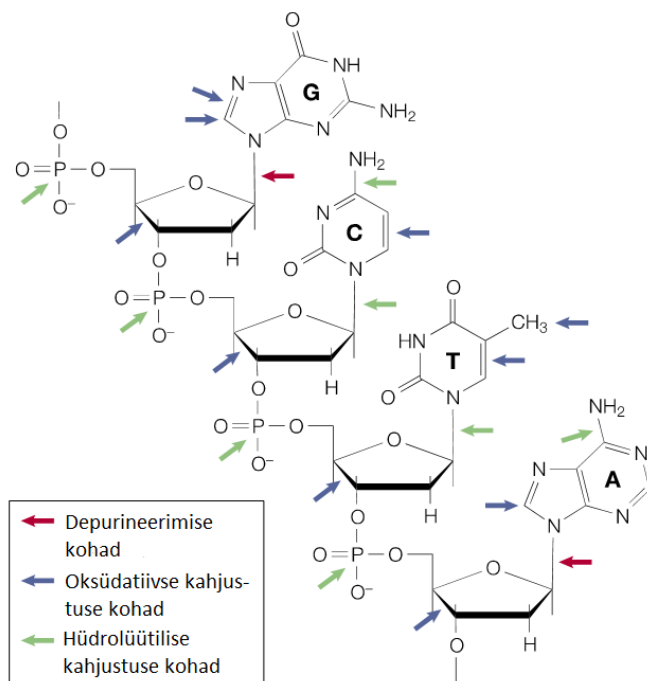
Samuti on sekveneeritud rida muumiatele kuuluvaid DNA järjestusi, vaatamata sellele et Egiptuse kuum kliima, enamike hauakambrite kõrge niiskus ning mumifitseerimisel kasutatavad ja DNA lagunemist soodustavad ained (eriti naatriumkarbonaat) vähendavad autentse DNA esinemise tõenäosust (Gilbert jt., 2003). Egiptuse muumia DNA analüüsi katse tegi esimesena Svante Pääbo 1985. aastal kasutades selleks molekulaarset kloonimist (Pääbo, 1985), mille tulemused on siiani vaieldavad. Alles väga hiljuti sai võimalikuks kõigile kvaliteedinõuetele vastav uuring, kus analüüsiti 151 erinevaist ajastutest pärinevat Egiptuse muumiat, millest 90-st õnnestus mtDNA eraldamine ja kolmest kogu genoomi informatsiooni

saamine. Samas näidati, et geneetiline materjal säilib ka pehmetes mumifitseeritud kudedes, kuid seal on amplifitseeritava DNA sisaldus keskmiselt 10 korda väiksem, kui luudes ja hammastes (Schuenemann jt., 2016).

Seega, kuigi DNAd võib eraldada erinevatest kehaosadest ja kudedest, kasutatakse põhiliselt kaltsifitseerunud kudesid – luid ja hambaid, millest omakorda suurimat rolli mängivad hambatsement ja kaljuluu.

1.2.2 DNA degradatsioon ja selle sõltuvus keskkonnatingimustest

DNA halb säilivus on iidsete genoomide uurimise peamiseks limiteerivaks teguriks. Kuna pärast organismi surma ei toimi enam rakus DNA reparatsiooni mehhanismid, on DNA eksponeeritud oksüdeerivatele ja hüdrolyütilistele kahjustustele, mille tekkimisel mängivad oma osa ka raku enda nukleasid. Peamisteks *post-mortem* ehk surmajärgselt tekkivateks DNA kahjustusteks on DNA lagunemine lühikesteks, kuni mõnesaja aluspaari pikkusteks fragmentideks, pürimidiinide ja suhkrujääkide modifikatsioonid ning lämmastikaluseta positsioonide ja ahelate- ja molekulidevaheliste ristsidemete teke (Pääbo, 1989, Lindahl, 1993). DNA põhilised keemilised kahjustused ja nende tekkekohad on näidatud joonisel 2. Vana DNA *post-mortem* modifikatsioonid raskendavad selle analüüsi: näiteks fragmentide väike pikkus ja kovalentsed sidemed raskendavad DNA amplifikatsiooni ning nukleotiidide modifikatsioonid põhjustavad valepaardumisi ja sekveneerimisvigu. Samas kasutatakse vana DNAle iseloomulikke muutusi selle autentsuse tõestamiseks, millest tuleb juttu edaspidi.



Joonis 2. Vana DNA põhilised *post-mortem* kahjustused. On näidatud DNA ühe ahela lühike lõik, milles on esindatud kõik neli nukleotiidi; kahjustuste kohad ja tüübid on näidatud värviliste nooltega (kohandatud Hofreiter jt., 2001 järgi).

Üheks olulisemaks teguriks, mis piirab vana DNA analüüsi standartsete meetodite abil, on selle lagunemine lühikesteks fragmentideks. Allentofti ja kaasautorite hinnangul on mtDNA lagunemise kiirus (fragmentatsioonitase) 5.5×10^{-6} nukleotiidi kohta aastas (vastav poolestusaeg on 521 aastat), kusjuures tuuma DNA laguneb vähemalt kaks korda kiiremini (Allentoft jt., 2012). Teiselt poolt aga on uuringuid, kus on näidatud, et kuigi DNA fragmenteeritus on vanadel proovidel kõrgem kui suhteliselt noortel, ei esine aja ja keskmise fragmendi pikkuse vahel selget korrelatsiooni (Sawyer jt., 2012).

DNA lagunemise kiirus sõltub paljudest surnukeha ümbritseva keskkonna füüsikalistest, keemilistest ja bioloogilistest teguritest, milleks on näiteks niiskus, mikroorganismide rohkus ja aktiivsus ning pH. Eelkõige on lagunemise kiirust määravateks teguriteks temperatuur ja niiskus. Soojus ja kõrge niiskustase soodustavad kiiret lagunemist samas kui külm ja kuiv keskkond pärsib mikroorganismide kasvu ja minimeerib keemilisi kahjustusi (Burger jt., 1999, Lindahl, 1993). Kiire kuivamine, külmumine või kõrge soolakontsentratsioon inaktiveerivad nukleaase nii, et nad ei kahjusta DNAd vahetult peale indiviidi surma (Nicholls, 2005). Lisaks akumulereuvad erinevad DNA kahjustused keskkonnatingimustest sõltuvalt erineva kiirusega: näiteks pakutakse, et igikeltsa tingimustes oleval DNAs tekivad ristsidemed kordades kiiremini kui katked (Hansen jt., 2006).

Vanimad analüüsitud proovid on eraldatud igikeltsast leitud jäänustest: näiteks praegu teadaolev vanim sekveneeritud DNA (560–780 tuhat aastat vana, 1.12-kordne kattuvus) on eraldatud igikeltsas olnud hobuseluust (Orlando jt., 2013). DNA säilimise seisukohast võiksid olla ideaalseks paigaks polaaraladel olevad jääkilbid, kus keskmine temperatuur on -50°C , kuid tasub mainida, et teoreetiliselt ei säili amplifitseeritav DNA isegi ideaalsetes tingimustes üle miljoni aasta (Willerslev ja Cooper, 2005). Kuigi mõõdukamas kliimas ei säili DNA nii hästi kui igikeltsas, sai tehnoloogiate täiustamisega võimalikuks ka väga vana DNA eraldamine madalatelt laiuskraadidelt: näiteks võib tuua 430 000 aastat vana Hispaaniast leitud hominiinide jäänustest eraldatud DNA sekveneerimist (Arsuaga jt., 2014).

Seega viitavad eelnevalt loetletud DNA surmajärgse degradatsiooni iseärasused sellele, et amplifitseeritavate DNA järjestuste sisaldus vanades kudedes on väga madal ning ka need esinevad seal märkimisväärselt fragmenteeritud ja modifitseeritud kujul, mis määrab edasise analüüsi spetsiifika.

1.2.3 Vana DNA eraldamine ja analüüsi spetsiifika

Töö vana DNAGA algab selle eraldamiseks sobivate kudede väljavalmimisest. Võimaluse korral valitakse hästi säilinud tahkeid kudesid ning suuri luid eelistatakse väikestele, kuna nendest õnnestub suurema tõenäosusega autentse DNA eraldamine.

DNA eraldamine iidsetest jäänustest algab kudede tükeldamisest või pulbriks hõõrumisest. Seejärel dekaltsifitseeritakse proovid inkubeerides neid etüleendiamiin-tetraäädikhappe lahuses (EDTA) 2-4 päeva jooksul ning puhastatakse mitme teise aine, näiteks proteinaas K abil, et eemaldada lahusest valke ja teisi ebavajalikke komponente. Sellele järgneb DNA sadestamine ja pesemine etanoolis, kuivatamine ja lahustamine Tris-EDTA (TE) puhvrts. DNA eraldamise protokollid erinevad sõltuvalt koest või organismist, eraldamise spetsiifika määrab ka eksemplari vanus ja säilivusaste (Rohland ja Hofreiter, 2007).

DNA eraldamisele järgneb selle molekulaargeneetiline analüüs: amplifitseerimine ja sekveneerimine. Vana DNA madal sisaldus ja degradeeritud iseloom piiravad selle uurimist tavaliste meetoditega, mis sobiksid elusate inimeste kudedest eraldatud DNA analüüsiks. Seetõttu on kogu vanade proovidega seotud uuringute ajaloo vältel pidevalt leiutatud ja täiustatud viise, kuidas suurendada vanade järjestuste kvaliteeti ja usaldusväärsust. Näiteks kasutatakse *post-mortem* modifikatsioonide hulga vähendamiseks uratsiil-DNA-glükosülaasi (UNG), mis eemaldab DNAST uratsiili, mis normaalselt seal ei esine, aga moodustub tsütosiini deamineerimise tulemusel (Hofreiter jt., 2001). Kõrge täpsusega DNA polümeraasid (*high-fidelity DNA polymerases*) teevad võrreldes “tavaliste” polümeraasidega vähem vigu suurendades niiviisi sekveneerimise efektiivsust. Samuti kasutatakse N-fenatsüül tiasoolium bromiidi (*N-phenacylthiazolium bromide*, N-PTB), mis lõikab ristsidemeid suhkrujääkide ja aminorühmade vahel, muutes seeläbi DNAd amplifitseeritavaks (Poinar jt., 1998).

1.2.3.1 Vana DNA analüüsi varased meetodid

Esimestes vana DNAGA seotud uuringutes kasutati molekulaarset kloonimist, et paljundada loomanahkadest ja inimmuumiatest eraldatud DNA fragmente (Higuchi jt., 1984, Pääbo S., 1985). Molekulaarne kloonimine on meetod, mis põhineb uuritava DNA paljundamisel elusorganismis, eelkõige bakteris *E. coli*. Analüüsitav DNA liidetakse teise DNA elemendiga, mida kutsutakse vektoriks. Vektori DNA sisaldab kõiki vajalikke elemente, et peremeesorganismis replitseeruda. Lisaks sisaldab vektor tavaliselt geeni, mille produkt tagab resistentsuse mingile antibiootikumile, et oleks võimalik positiivselt selekteerida vektorit sisaldavaid peremeesrakke. Analüüsitava fragmendi sisestamiseks vektorisse kasutatakse tavaliselt DNAd lõikavat ensüümi – sait-spetsiifilisi DNA endonukleaase ehk restriktase – ja kaheaahelalisi katkeid kokkuliitvat ensüümi – DNA ligaasi. Seejärel sisestatakse analüüsitavat

järjestust sisaldav plasmiid peremeesrakku ning ta replitseerub selles rakuliste mehhanismide abil [6].

Antud meetodi rakendamise näiteks võib tuua 1989. aasta uuringu, kus analüüsiti nelja loomaliigi jäänustest eraldatud DNAd; molekulaarse kloonimise teel paljundatud DNA fragmendid olid keskmiselt 90 ap pikad (Pääbo S., 1989). DNA analüüsil molekulaarse kloonimise abil on mitu puudust. Meetod vajab palju aega, mis kulub nii piisava arvu fragmentide saamiseks ehk bakterite paljunemiseks, kui ka analüüsitava DNA järgneva puhastamiseks rakumassist. Lisaks on ligeerimise ja transformeerimise madal efektiivsus materjali märkimisväärsete kadude põhjuseks. Meetod võimaldab vaid üksikute suure koopiaarvuga (põhiliselt mtDNA) lookuste analüüsi ning eraldatava vana DNA väikese hulga tõttu on praktiliselt võimatu ühesuguseid järjestusi kandvate bakteriaalsete kloonide isoleerimine, mistõttu ei ole tulemused hästi korratavad (Pääbo jt., 2004). Tasub mainida, et lisaks molekulaarse kloonimise nõrkustele puudusid vana DNA uuringute algusajal ka usaldusväärsed meetodid DNA kahjustatuse ja saastatuse hindamiseks.

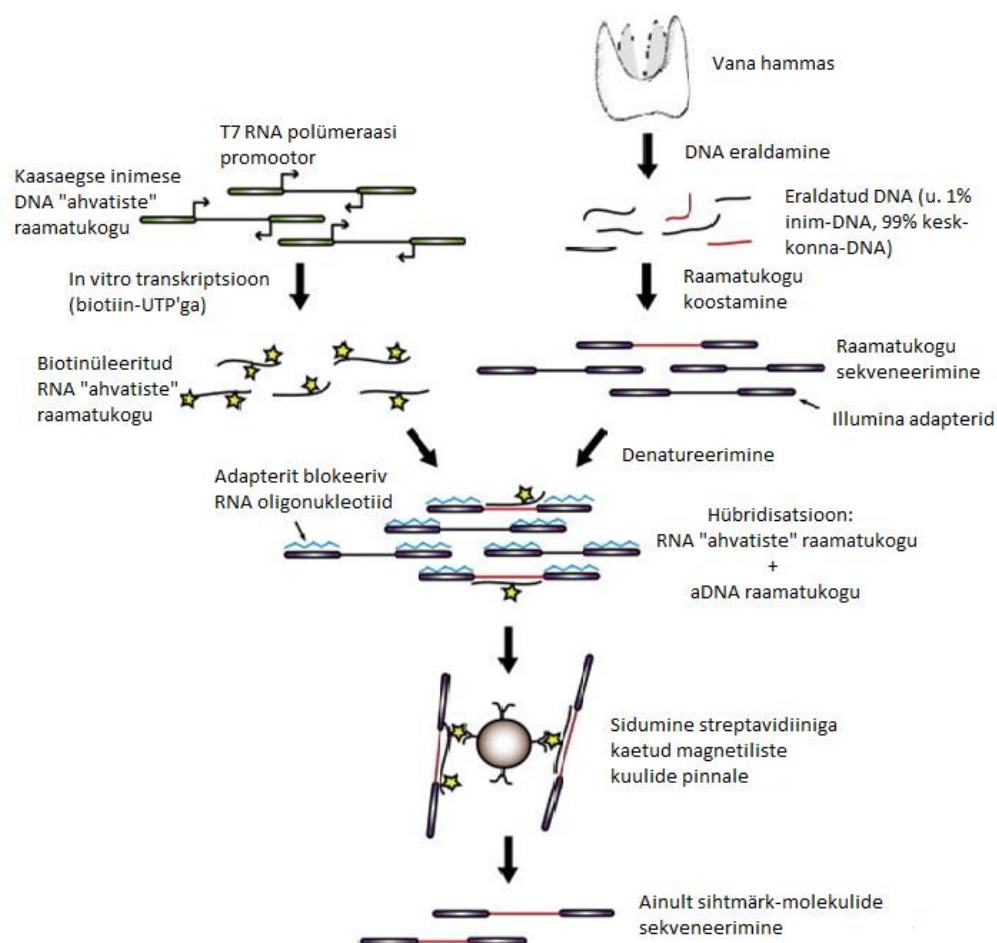
Polümeraasi ahelreaktsiooni (PCR) laialdase kasutamise algusega muutus vana DNA analüüs palju lihtsamaks, kuna PCR võimaldab isegi üksikutest molekulidest piiramatult arvu koopiate saamist lühikese aja jooksul. Kuid ka sellel meetodil on omad miinused: selle rakendamiseks peab teadma amplifitseeritava DNA järjestust, et konstrueerida sobivaid primereid; DNA polümeraas teeb ja paljundab vigu; PCR on võimalik vaid juhul kui amplifitseeritav fragment on piisavalt pikk, et selle informatiivse järjestusega võiks seostuda kaks praimerit (Dabney jt., 2013). Samas on pikki fragmente vanades proovides vähe ja kõige rohkem on väga lühikesi molekule, mistõttu sisaldab just see fraktsioon kõige rohkem informatsiooni (Allentoft jt., 2012, Orlando jt., 2013). Lisaks suurenes PCR kasutamisega ka vanade proovide tundlikkus kaasaegse DNAGA saastuse suhtes, kuna kõrgema kvaliteedi tõttu on nende amplifikatsioon palju efektiivsem kui degradeeritud ja modifitseeritud vanal DNAl (Willerslev ja Cooper, 2005).

1.2.3.2 Vana DNA molekulaargeneetiline analüüs tänapäeval

Tänapäeval on vana DNA järjestuste saamise põhiliseks viisiks täisgenoomne sekveneerimine NGS meetodite abil, mis võimaldab tugevalt fragmenteeritud DNA efektiivset sekveneerimist ning seeläbi informatsiooni kadude minimeerimist. Nüüdisajal kasutatakse kogu genoomi sekveneerimiseks eeskätt firma Illumina toodetud platvorme. Molekulaargeneetiline analüüs algab sekveneerimisraamatukogu koostamisest [7]: vana DNA fragmentide otsad parandatakse ja nende otstes liidetakse adapterid, millele komplementaarsed oligonukleotiidid on immobiliseeritud tahkele kandjale. Adapterite kaudu

seostunud järjestused amplifitseeritakse sildamplifikatsiooni abil (*bridge amplification*), mille tulemusel on pind kaetud identsete molekulide klastritega. Seejärel toimub sekveneerimine (*sequencing by synthesis*), kus kasutatakse fluorestsentselt märgistatud nukleotiide ning signaali lugemine toimub iga sekveneerimistsükli järel [8].

NGS meetodite peamiseks eeliseks on molekulide kogu pikkuse sekveneerimine ning see, et protsess ei ole piiratud praimerite spetsiifiliste seostumissaitidega (Skoglund jt., 2014). Samuti võib selle tehnoloogia abil amplifitseerida ka väga lühikesi DNA fragmente (<50 ap), mis on suureks plussiks kõrgelt fragmenteeritud vana DNA analüüsil (Dabney jt., 2013) ja võimaldas uuritavat ajaakent lükata suurusjärgu võrra varasemasse aega (Orlando jt., 2015).



Joonis 3. Vana DNA proovi rikastamine WISC meetodil. Luuakse kaasaegse inimese DNA baasil DNA “ahvatiste” raamatukogu, mida transkribeeritakse biotinüleerides *in vitro*, seejärel hübridiseeritakse saadud RNA-“ahvatiseid” arheoloogilisest proovist eraldatud DNA raamatukoguga ning seotakse biotinüleeritud nukleotiide sisaldavaid DNA-RNA hübride streptavidiiniga kaetud tahke kandja pinnale, millele järgneb seotud DNA lõikude sekveneerimine (kohandatud Carpenter jt., 2013 järgi).

Vanadest kudedest eraldatava DNA kogus on tavaliselt väga madal, mis raskendab vana DNA analüüsi. Olukorra parandamiseks on välja töötatud meetodid, mis on suunatud proovide rikastamisele endogeense DNAGA: seostamine tahkele kandjale (tavaliselt klaasist

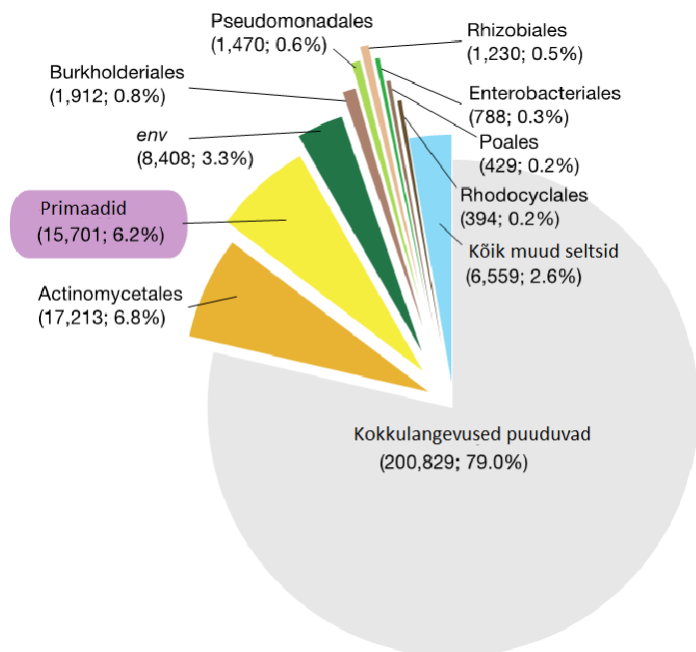
mikrokandjad) või sidumine lahuses (Bainbridge jt., 2010). Siduda ja paljundada võib nii mtDNA ja kromosoomide üksikuid regioone kui ka eksoomi või tervet genoomi. Lahuses kinnipüüdmise näiteks võib tuua terve genoomi kinnipüüdmine lahusest (*whole-genome in-solution capture*, WISC), mille abil saab kätte kogu proovides sisalduva informatsiooni ja mis seetõttu sobib kõige paremini degradeeritud vana DNA analüüsiks (joonis 3). Selleks kasutatakse biotinüleeritud RNA-oligonukleotiide (“ahvatisi”), mis on konstrueeritud kaasaegse inimese genoomi baasil ning katavad kogu genoomi. Lahuses toimub hübriidsatsioon ja fragmentide sidumine streptavidiiniga kaetud kuulikeste pinnale. Seondumata DNA, mis pärineb teistelt organismidelt, pestakse välja, samuti eemaldatakse kogu RNA, nii et kuulide pinnale jääb vaid analüüsimisele kuuluv inim-DNA, mida seejärel amplifitseeritakse ja sekveneeritakse (Carpenter jt., 2013).

Kirjeldatud meetod ja teised sarnased meetodikad hõlbustavad märkimisväärselt vana DNA analüüsi: näiteks oli ülalmainitud töös (Carpenter jt., 2013) enne rikastamist 1.2% järjestustest määratud inimesele kuuluvaks ja pärast rikastamist tõusis see protsent umbes 60-ni. Tänu sellele väheneb madala endogeense DNA kontsentratsiooniga proovide sekveneerimise keerukus ja hind ning suhteliselt lühikese aja jooksul on võimalik sekveneerida palju proove.

1.3 Vana DNA saastus

Vana DNA saastus on uuritava organismi DNA segunemine eksogeensete, s.t. sellele organismile mittekuuluvate DNA järjestustega, mis raskendavad analüüsi ja võivad viia ekslike tulemusteni. Iidsetest jäänustest eraldatud DNA kujutab endast alati endogeense ja keskkondliku (s.t. bakteritele, seentele, taimedele ja muudele organismidele kuuluva) DNA segu (nt. Green jt., 2006). Tavaliselt on endogeense ehk autentse DNA sisaldus kogu eraldatud DNA taustal väga madal. Tihtipeale moodustavad autentsed järjestused alla 1% kogu DNA fraktsioonist (García-Garcera jt., 2011, Fu jt., 2012) ning vaid mõnedel juhtudel on see protsent kõrge: näiteks igikeltsa tingimustes (mis on ebasoodsad kolonisatsiooniks mikroorganismide poolt) olnud mammuti DNA analüüsil määrati mammutile kuuluvaks ligi pool kõigist järjestustest (Poinar jt., 2006).

Allpool on toodud joonis (joonis 4), mis illustreerib neandertaallase luuproovist eraldatud DNA koostist. Võib näha, et suurema osa DNA jaoks oli selle allikaks olnud organismi taksonoomilise kuuluvuse määramine võimatu, kuna andmebaasist ei leitud piisavalt sarnaseid järjestusi, ja primaatidele kuulus 6.2% järjestustest.



Joonis 4. Näide neandertaallase luuproovist eraldatud DNA taksonoomilisest koostisest (kohandatud Green jt., 2006 järgi). Primaatidele kuuluv DNA on eraldi välja toodud lillal taustal.

Inimjäänustest eraldatava DNAGA seotud töödes on kõige ohtlikumaks saastajaks kaasaegse inimese DNA. Kui näiteks väljasurnud mammuti uurimisel on üsna väike tõenäosus, et proovidesse satub kaasaegse elevandi DNA, siis inimesega seotud uuringute korral on potentsiaalsed saasteallikad pidevalt proovide vahetus läheduses. Kuna tänapäeval elava inimese DNA ei ole kergesti eristatav iga teise AMI DNAST sõltumata sellest, kui ammu aega tagasi viimane elas (Sawyer jt., 2012), võib see viia ekslike järeldusteni. Kui räägitakse vana DNA proovi saastusest, siis mõeldakse selle all just saastust kaasaegse inimese DNAGA ning selles tähenduses kasutatakse seda mõistet edaspidi ka käesolevas töös. Vaatamata arvukatele kaitseabinõudele, mida rakendatakse vähendamaks eksogeense DNA sattumist iidsetesse proovidesse, jääb saastus inimese paleogenoomika üheks tõsisemaks probleemiks.

1.3.1 Saastuse peamised allikad ja põhjused

DNA hakkab lagunema kohe peale organismi surma ning selle degradatsioonimäär suureneb ajaga kuni see muutub analüüsiks kõlbmatuks. Vana DNA amplifitseerimine on raskendatud võrreldes kaasaegse parema kvaliteediga DNAGA, mistõttu võib isegi väike kogus segusse sattunud kaasaegset DNAd osutada saastuse seisukohast ohtlikuks.

Kaasaegse inimese DNA võib sattuda proovidesse igas tööetapis: väljakaevamise ajal, proovide transportimisel, säilitamisel ja analüüsimisel. Väljakaevamise vältel, kui luumaterjal kaevatakse välja, pestakse, sorteeritakse ja märgistatakse, võivad seda saastada proovidega töötavate inimeste (arheoloogide, antropoloogide, paleontoloogide jne) rakud. Rakud võivad

pärineda kõikvõimalikest kehaeritistest: nii naha- ja juukserakkudest kui ka süljetilgakestest ja higist. Kuigi saastumine võib toimuda igas tööetapis, on mõned etapid kriitilisemad kui teised. Näiteks Sampietro ja kaasautorite uuringus oli üle 17% hammaste analüüsi tulemusel saadud DNA järjestustest identifitseeritud kui saastajad ning sagedamini pärinesid nad inimestelt, kes tegelesid proovide väljakaevamise ja pesemisega kui antropoloogidelt ja molekulaargeneetilist analüüsi läbiviinud teadlastelt (Sampietro jt., 2006). Niisiis osutus saastuse seisukohalt kõige ohtlikumaks etapp, mis järgnes vahetult arheoloogilise materjali väljakaevamisele.

Ka luu ehitus mängib olulist rolli selle reostamisel. Luukude eristub muudest kudedest oma poorsuse poolest: elusa organismi luu mahust moodustab umbes 8% õhk ja ammu surnud organismi luud koosnevad 40-50% õhust, mistõttu on üsna tõenäoline, et sinna satuvad igasugused sellega kokku puutuvad eksogeensed ained, sealhulgas ka kaasaegne DNA (Nicholls, 2005). Seega sõltub luumaterjali saastatuse määr enne laborisse sattumist ka selle säilivusastmest: mida tugevamini on kude degradeerunud ja lagunened, seda vastuvõtlikum on see saastusele (Gilbert jt., 2006).

Mitte vähem ohtlik on saastus, mis satub proovidesse laboritingimustes. Selle põhjuseks võib olla nii degradeeritud materjaliga töötamise juhiste eiramine, kui ka saastunud reaktiivide kasutamine. Selle vältimiseks rakendatakse erinevaid kaitseabinõusid, mida on lähemalt kirjeldatud alapeatükis 1.4.2.

Lisaks võivad tekkida vead andmeanalüüsil kui näiteks mtDNA sekveneerimisel saadud järjestused *in silico* valesti haplotüüpidesse määrata.

1.3.2 Vana DNAGA töötamise peamised reeglid ja kriteeriumid

Vana DNAGA töötamise reeglid sõnastati ja süstematiseeriti 2000. aastal (Cooper ja Poinar, 2000), NGSi kasutuselevõtuga on need veidi modifitseeritud ja täiendatud (Knapp jt., 2012). Peamiste punktidenä vana DNAGA töötamisel võib välja tuua järgmised nõuded:

- 1) füüsiliselt isoleeritud tööruum;
- 2) kontrolleraldamised ja kontrollamplifikatsioon;
- 3) tulemuste korratavus nii ühe ja sama, kui ka erinevate eraldamiste jaoks ühest proovist;
- 4) tulemuste korratavus erinevates laborites;
- 5) biokeemiliste kahjustuste hinnang (*post-mortem* modifikatsioonid jms);
- 6) kvantitatiivne hinnang (kasutatakse reaajas kvantitatiivset PCR-reaktsiooni, *real-time quantitative PCR*, RT-qPCR (Morin jt., 2001)) ehk algse matriitsjärjestuste arvu määramine: kui see on madal (alla 1000), võib see viitada sporaadilisele saastusele.

Kaasajal töötatakse vana DNAGA eranditult spetsiaalselt selleks ehitatud ja seadmestatud laborites (Fulton, 2011): need on eraldiasuvad hooned või ruumid, millel ei ole füüsilist kontakti laboritega, kus analüüsitakse kaasaegset DNAd. Vana DNA laborites on mitu osakonda erinevate tööetappide jaoks. Neid laboreid ventileeritakse rakendades positiivset õhurõhku, kus sissetulevat õhku filtreeritakse, ning kõiki tööpindu kiiritakse UV-kiirgusega ja töödeldakse DNAd lagundavate ainetega. Personal töötab spetsiaalses kaitseriietuses (jalatsikatted, laborikitlid, kindad, kaitsemaskid ja juuksekaite) ning liigub tööpäeva vältel vaid puhasruumist kaasaegsete laborite suunas vältimaks saastust eksogeense DNAGA. Kogu laborisse saabuv materjal puhastatakse ja töödeldakse kloori sisaldavate lahustega, mis lagundavad pinnal oleva saastava DNA (Kemp ja Smith, 2005) ning seejärel töödeldakse UV-kiirgusega vähemalt 30 minuti jooksul. Samamoodi töödeldakse ka kõikide reaktiivide, seadmete ja lahuste pealispindu ja mahuteid (Pinhasi jt., 2015). Eraldamise ja molekulaargeneetilise analüüsi etapid viiakse läbi eraldi ruumides ning kõik etapid viiakse läbi tõmbe- või laminaarkapi all.

1.3.3 Saastuse tuvastamise viisid

Autentifikatsioon on protseduur, mille käigus kinnitatakse DNA kuuluvust proovile, millest see eraldati. Autentifikatsioon on vana DNA uuringute kriitilise tähtsusega etapp, kuna sellest sõltub kogu uuringu tulemuste tõesus. Eksisteerib rida protseduure, mille abil on võimalik veenduda iidsetest jäänustest eraldatud DNA autentsuses.

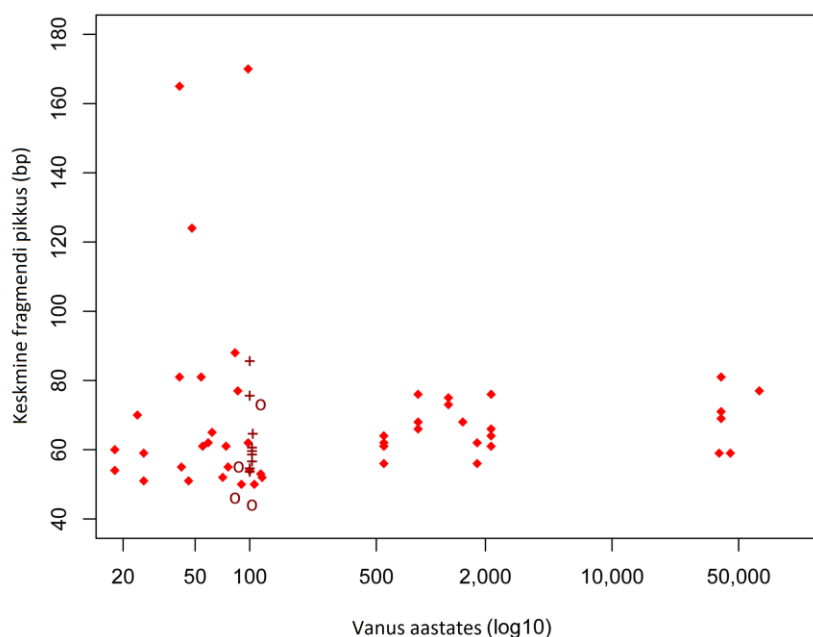
1.3.3.1 Kontrollproovid

Vastavalt vana DNAGA töötamise kriteeriumitele (Cooper ja Poinar, 2000) on soovitatav iga proovi jaoks mitme järjestuse saamine. Esiteks võib eraldada DNAd paralleelselt mitmes laboris, mis välistab laborispetsiifilise saastuse. Teiseks võib teostada ka ühe labori piires üle ühe eraldamise iga proovi kohta, näiteks jäänuste erinevatest kehaosadest. Lisaks võib sekveneerida igat amplifikatsiooniproducti mitmes korduses. Nendel viisidel saadavad järjestused peavad olema identsed.

Negatiivne kontroll on “tühja” ehk DNA-vaba proovi analüüsimine paralleelselt proovidega, kuhu uuritav materjal on lisatud. Negatiivne kontroll aitab tuvastada eelkõige reagentide ja muude lahuste saastust. Kuna ülalkirjeldatud identsuse kontroll on töömahukas ja kallis, kasutatakse tänapäeval enamasti negatiivset kontrolli ja vanale DNAle iseloomulike tunnuste esinemise määramist (1.3.3.2 – 1.3.3.3).

1.3.3.2 Sekveneerimisfragmentide pikkuse analüüs

Rida DNA autentsuse kontrollimise viise baseerub degradeerunud vana DNA molekulide iseloomulike karakteristikute jälgimisel. Veendumaks, et tegemist on tõesti vana molekuliga, võib kasutada DNA omadust aja jooksul laguneda (joonis 5). Suurem osa vanast DNAST on lagunenenud alla 100 ap pikkusteks fragmentideks (Hofreiter jt., 2001), ja enamus fragmentidest võivad olla isegi alla 30 ap pikad (Meyer jt., 2012). Seega võib teha sekveneeritud fragmentide pikkuste jaotuse järgi järeldusi DNA lagunemisastme kohta ning sellest tulenevalt ka DNA autentsuse kohta.



Joonis 5. MtDNA keskmine fragmendi pikkus erineva vanusega imetajajäänuste jaoks. Keskmine fragmendi pikkus varieerub 44-st 170 ap-ni. Kolm mtDNAd, millel see ületab 100 ap, kuuluvad alla 100 aasta vanadele jäänustele. Ristidega on märgistatud ahvijäänuste proovid, mida oli teadaolevalt “röstitud” tulel, maetud ja peale väljakaevamist töödeldud muuseumis polüvinüülatsetaadil põhineva puiduliimiga. Ringidega on märgistatud jäänuste proovid, mida muuseumis töödeldi väävelhappel põhineva säilitusainega. Rombidega tähistatud jäänuste proovid ei läbinud erilist töötlust (kohandatud Sawyer jt., 2012 järgi).

1.3.3.3 Surmajärgsete modifikatsioonide analüüs

Peale organismi surma DNAsse akumuliseeruvate modifikatsioonide iseloomu ja jaotuse analüüs võib ühtlasi olla autentsuse kontrolliks. On näidatud, et DNA katkeb enamasti puriinide, s.t. adeniini ja guaniini ees, nii et vana DNA fragmentide 5'-otste läheduses on jälgitav nende suurem osakaal, kui fragmentide keskel, kusjuures noorema DNA freagmentidel on otste lähedal rohkem adeniini ja vanema omadel guaniini. Lisaks suureneb fragmentide 5'-otstele lähenedes tsütosiini deamineerimistase türosiinini ning on näidatud, et taoliste modifikatsioonide osakaal suureneb aja jooksul (Sawyer jt., 2012). 2007. aasta uuringus

näitasid Brotherton ja kolleegid, et nende analüüsitud erineva liigilise kuuluvuse, päritoluregiooni ja vanusega vana DNA järjestustes moodustasid tsütosiini deamineerimised üle 90% kõikidest nukleotiidsetest asendustest (Brotherton jt., 2007). Jälgides erinevate nukleotiidide osakaalusid DNA fragmentide otste läheduses, saab niimoodi tuvastada erineva vanusega DNA proove.

1.3.3.4 Vana DNA võrdlemine kaasaegse DNAGA

Vaatamata sellele, et vanal DNAl on iseloomulikke modifikatsioone, võib teatud juhtudel ka kaasaegne DNA olla kahjustatud sarnasel moel. Näiteks näitasid 2006. aastal Sampietro ja kolleegid, et üle 10 aasta vanad saastajad ei ole kahjustuste järgi eristatavad uuritavast DNAST (Sampietro jt., 2006). Lisaks võib ka kloori sisaldavate komponentidega töötlus põhjustada kaasaegses DNAs vana DNA omadele sarnaseid kahjustusi (Garcia-Garcera jt., 2011). Seetõttu osutub väga tähtsaks saastuse detekteerimise võimaluseks vana DNA järjestuste võrdlus potentsiaalsete saasteallikate DNA järjestustega (nt. Caramelli jt., 2008).

Esiteks võib vaadata vana DNA järjestusi üldise kaasaegse DNA mitmekesisuse taustal. Kui vana DNA tuleneb populatsioonist, kus esinesid alleelid, mis on kaasaegsetel inimestel väga haruldased, võib neid erinevusi kasutada vana ja kaasaegse DNA eristamiseks üksteisest (nt. Ermini jt., 2008). Eristamiseks vana DNAd vaadatakse eelkõige “diagnostilisi” positsioone mtDNAs, milles uuritav populatsioon erineb kõigist kaasaegsetest inimestest. Mittevastavused vanade ja kaasaegsete järjestuste vahel peavad olema suurimad aladel, mis on polümorfseid tänapäevastes populatsioonides (Rasmussen jt., 2011). See aitab ka sporaadilise saastuse tuvastamisel PCR tulemusel saadud järjestuste järgi: kui lisaks autentsele oli kasutatud ka väike hulk kaasaegseid järjestusi, võivad nad ilmnedagi topelt-järjestustena polümorfsetes positsioonides (Skoglund jt., 2013). Teiseks meetodiks on indiviidi bioloogilises soo määramine võrreldes X ja Y kromosoomi fragmentide hulka vana DNA proovis, misjärel võib hinnata saastatust vastassugupoolele kuuluva DNAGA. Lisaks kasutatakse DNA autentifitseerimiseks ka statistilisi demograafilisi analüüse, näiteks suurima tõenäosuse hindamine (*maximum likelihood estimation*) põhineb asjaolul, et varem elanud hominiinide (näiteks neandertaallaste) genoomis esinevad kõrge sagedusega eellasalleelid, mis esinevad ka šimpansidel, ja kaasaegsel inimesel esinevad sagedamini hilisemal ajal tekkinud alleelid, mis ei lineariseeru šimpanside DNA järjestusega (Green jt., 2010).

Kuid nagu oli juba ülalpool mainitud, osutuvad kõige tihedamini vanade proovide saasteallikaks vana luumaterjali ja sellest eraldatud DNAGA töötavad inimesed. Seda asjaolu kasutatakse saastuse tuvastamiseks, sekveneerides potentsiaalsete saasteallikate DNAd ja võrreldes saadud järjestusi vanadega, s.t. vaadates individuaalseid erinevusi konkreetsete

indiviidide DNA järjestuste vahel. Tavaliselt kasutatakse võrdlusanalüüsiks mtDNAd, mis selle suurema sisalduse ja polümorfismi tõttu sobib kõige paremini saastuse tuvastamiseks (Sampietro jt., 2006).

Kokkuvõtteks võib öelda, et vigade vältimiseks on vaja kompleksset analüüsi, mis ei piirdu vaid ühega ülaltoodud meetoditest ning kasutab nii vana DNA molekulaarseid iseärasusi ja vanadele populatsioonidele iseloomulikke alleelseid mustreid kui ka võimalike saastajatega võrdlemist.

1.4 Saastuse tagajärjed

Vanade proovide saastus on põhjustanud rea valepositiivseid tulemusi ja nendel baseeruvaid publikatsioone. Eriti palju mitte-autentsetel tulemustel põhinevaid artikleid on ilmunud vana DNA uuringute algetappidel, kui vana DNA töö kõiki raskusi ja piiranguid veel ei teatud ning ei olnud veel kindlaid reegleid ja kriteeriume, mis aitaksid vältida saastust ja kontrollida järjestuste autentsust.

Vaid paarkümmend aastat peale vana DNA uuringute algust määrati DNA vanusepiir, mida määratlevad selle molekulaarkeemilised omadused: isegi ideaalsetes tingimustes ei säili amplifitseeritav DNA üle miljoni aasta (Willerslev ja Cooper, 2005). Kuid paljude varasemate uuringute autorid arvasid, et said miljoneid aastaid vanu DNA järjestusi. Näiteks väitsid Woodward ja kaasautorid, et neil õnnestus sekveneneerida 80 miljoni aastase dinosauruse luudest eraldatud DNA (Woodward jt., 1994). Vreeland ja kolleegid kirjutasid oma artiklis, et kultiveerisid soolakristallis 250 miljonit aastat veetnud bakterit (Vreeland jt., 2000). Golenberg'i ja kaasautorite töös oli kirjeldatud 820 ap pikkune DNA fragment, mis oli väidetavalt eraldatud magnoolia kloroplastist vanusega 17-20 miljonit aastat (Golenberg jt., 1990); Cano ja kolleegide artiklis – 25-40 miljonit aastat merevaigus veetnud mesilase DNA (Cano jt., 1992).

Mõnele sellistest “avastustest” järgnesid neid ümberlukkavad tööd (nt. Walden ja Robertson, 1997). Näiteks taas analüüsides DNAd, mis pidi kuuluma dinosaurusele (Woodward jt., 1994) avastati, et fülogeneetiliselt on see üllatavalt lähedane inimesele järjestusele (nt. Hedges ja Schweitzer, 1995). Mõnede uuringute puhul ei ole aga “vana” järjestuse vale fülogeneetilise kuuluvuse tõestamine nii lihtne nagu ülalkirjeldatud juhtudel. Näiteks juhul kui analüüsitakse bakteriaalset DNAd, mida eraldati miljoneid aastaid vanadest soolakristallidest, ei ole puhtalt järjestuse põhjal võimalik väita, et tegemist on kaasaegse, mitte muistse bakteriaalse DNAGA. See-eest tehakse uuringuid tõestamaks taoliste tulemuste saamise võimalust. Näiteks panid Austin ja kolleegid DNA säilimise merivaigus miljoneid aastaid veetnud putukates kahtluse alla. Teadlased viisid läbi iseseisva uuringu, kus üritasid eraldada

DNAd 15-st niisugusest putukast ning neil ei õnnestunud mitte ühegi autentse DNA lõigu eraldamine (Austin jt., 1997).

Ka mõnede inimesega seotud uuringute tulemused on osutunud küsitavateks ja kontrollimist vajavateks. Näiteks 1994. aastal saadi Egiptuse muumiast “ebatavaline” mtDNA järjestus, mis erines märkimisväärselt mistahes teisest teadaolevast inim-mtDNAst (Hänni jt., 1994); kahe kromanjoonlase mtDNA Caramelli ja kolleegide uuringus on kahtlaselt sarnane kaasaegse inimese mtDNAlle (Caramelli jt., 2003). Rida ekslikke artikleid on seotud ka tuuma DNAGA, mille vähene hulk ja analüüsimise keerukus teevad veelgi kahtlasemaks ja kontrollimist vajavaks autentsete järjestuste saamise. Näiteks sekveneerisid Beraud-Colomb ja kolleegid suhteliselt pika fragmendi (380 ap) tuuma DNAd, mis oli eraldatud väidetavalt 12 000 aastastest inimluudest (Beraud-Colomb jt., 1995) – see on pigem saastuse tulemus, kui autentne järjestus. 1985. aastal uuris Svante Pääbo 23 muumiat molekulaarse kloonimise teel, kuid leidis üllatavalt vähe surmajärgseid modifikatsioone (Pääbo S., 1985). Nüüd on üldtunnistatud nende tulemuste ekslikkus (Shapiro ja Hofreiter, 2014).

Praeguseks ajaks on töötatud välja analüütilised meetodid, mis võimaldavad vanade proovide saastuse määra hindamist. Näiteks käesolevas töös käsitletavate vana DNA proovide jaoks kasutati saastustesti, mis põhineb proovi mtDNA konsensusjärjestuse koostamisel lühikestest sekveneerimisega määratud järjestustest, misjärel võrreldakse antud proovi lühikesi järjestusi nii selle konsensusjärjestuse kui ka ~300 mtDNA järjestusega andmebaasist. Saastusprotsent arvutatakse järgnevalt: (proovi lühikeste järjestuste kokkulangevused andmebaasis olevate järjestustega) / (proovi lühikeste järjestuste kokkulangevused proovi konsensusjärjestusega). Erinevate meetoditega ja erinevat (mt või tuuma) DNAd kasutades arvutatud saastuse protsent tuuakse välja artiklites: näiteks Prüfer’i ja kaasautorite uuringus oli neandertaallase DNA saastatud kaasaegse inimese DNAGA umbes 1% osas (Prüfer jt., 2014).

Kokkuvõtteks võib öelda, et vana DNA analüüsimetodiite (molekulaarse kloonimise) eripäradest ja kindlate väljatöötatud saastuse vältimise tööreeglite puudumisest tingitult on osad varased vana DNA-alased tööd ebausaldusväärsed. Kuid isegi kaasaegsetes tingimustes on saastusega arvestamine vältimatu ja sugugi mitte kõik ettevaatusabinõud ei pruugi selle vastu olla efektiivsed. Sellepärast on alati vajalik järjestuste autentsuse hoolikas kontrollimine. See puudutab eriti inimese DNAd, kuna kõiki proovidega kokkupuutuvaid inimesi peab käsitlema potentsiaalsete saasteallikatena.

2 Eksperimentaalsosa

2.1 Töö eesmärgid

Töö eesmärgiks oli vana DNA puhtuse kontrollimine. Selleks võrreldi keskaegse Eesti territooriumi elanikkonna proove arheoloogilise materjaliga kokku puutunud inimeste proovidega, kuna nende DNA võis suure tõenäosusega osutuda vanade proovide saastajaks. Eesmärgi saavutamiseks püstitati järgmised ülesanded, mis viidi läbi töö autori poolt:

- 1) Eraldada arheoloogide DNA kahest erinevast bioloogilisest materjalist;
- 2) Viia läbi DNA proovide kvalitatiivne ja kvantitatiivne hinnang;
- 3) Määrata mtDNA haplogrupid sekveneerides kogu mtDNA genoomid või sekveneerides HVS-regioone ja genotüpeerides kodeeriva regiooni informatiivseid positsioone;
- 4) Teostada vana DNA ja arheoloogide DNA võrdlusanalüüs.

2.2 Materjalid ja meetodika

2.2.1 Kasutatud proovid ja üldine töö käik

Käesolevas töös analüüsitud DNA proovid pärinevad Eesti arheoloogide rühmalt (10 inimest). Kontrolliks on analüüsitud kaks käesoleva töö autori proovi (proovid 11 ja 12), kuna DNA analüüsiga tegelevates laborites on see soovitatav kõikide töötajate jaoks. Nende järjestuste identsus näitab töö korrektsust.

Proovide analüüs algas DNA eraldamisest, millele järgnes kontsentratsiooni ja puhtuse mõõtmine ning vajadusel proovide kontsentreerimine ja puhastamine. Molekulaargeneetiline analüüs seisnes proovide amplifitseerimises ja sekveneerimises. Sellele järgnes saadud järjestuste analüüs spetsiaalsete arvutiprogrammide ja andmebaaside abil.

2.2.2 DNA eraldamine

Bioloogiliseks materjaliks kasutati põseepiteeli kaapmeid ja süljeproove. Nende kahe materjalitüübi jaoks kasutati erinevaid DNA eraldamise meetodikaid. Mõlemad meetodid põhinevad DNA seondumisel ränioksiidi sisaldavale filtrile kõrge soola kontsentratsiooni juuresolekul ning pesemisel 70% etanooli sisaldava lahusega.

2.2.2.1 DNA eraldamine põseepiteeli kaapmest

DNA eraldamiseks kasutati protokollfi firma Qiagen raamatust QIAamp DNA investigation Handbook 06/2012 (*Isolation of Total DNA from Surface and Buccal Swabs*). Kasutatud vahendid: 2 ml tuubid, QIAamp MinElute kolonnid, 1.5 ml tuubid. DNA eraldamine koosnes järgmistest etappidest:

1. Rakumembraanide lõhustamine ja valkude hüdrolüüs: sellel etapil lõhutakse proovis olevate rakkude membraanid kahe lüüsipuhvriga (ATL ja AL) ja lagundatakse valke kasutades proteinaas K lahust.

Selleks asetatati uuritavad proovid 2 ml tuubidesse, lisati nendele 400 µl ATL puhvrit ja 20 µl proteinaas K lahust (kontsentratsiooniga 20 mg/ml), segati vortex-segajal 10 s ning asetati proovid 1 tunniks termomikserisse (Thermo-Shaker TS-100 (Biosan)) 56°C, 900 rpm (*revolutions per minute*, pööret minutis) juurde. Seejärel lisati 400 µl AL puhvrit, segati vortex-segajal 15 s ning inkubeeriti proove 10 min termomikseris 70°C, 900 rpm juures.

2. DNA seondumine filtrile: segu asetati ränioksiidist filtriga kolonnile ja tsentrifuugiti 1 min 8000 rpm juures.

3. Filtri pesemine: DNA puhastamiseks lüsaadi muudest komponentidest pesti filtreid kasutades kaht pesupuhvrit (AW1 ja AW2) ja etanooli lahust. Proovidele lisati järjestikku 500 µl AW1 puhvrit, 700 µl AW2 puhvrit ja 700 µl 96% EtOH. Pärast iga komponendi lisamist tsentrifuugiti proove 1 min 8000 rpm juures ja vahetati tuube, kuhu kogunes eemaldatavaid lüsaadi komponente sisaldav eluaat. Seejärel tsentrifuugiti proove 3 min 13 400 rpm juures, et saada lahti etanooli jääkidest ning asetati proovid 3 minutiks termostaati 56°C juurde, et filtrimembraan kuivaks täielikult.

4. DNA elueerimine: sellel etapil eraldub DNA filtrilt ning see pestakse läbi filtri kogumistuubi. Selleks asetati filtrile 100 µl destilleeritud vett, hoiti proove 5 min toatemperatuuril ning tsentrifuugiti 1 min 13 400 rpm juures.

2.2.2.2 DNA eraldamine süljeproovidest

DNA eraldamiseks kasutati SalivaGene DNA Kit 0711 (Stratec Biomedical Systems). DNA eraldati süljeproovidest kahes korduses: esimesel korral kasutati elueerimiseks vett ja teisel korral elueerimispuhvrit. DNA eraldamisel kasutatud vahendid: 1.5 ml tuubid, RTA Spin Filter Set (Stratec Molecular GmbH). DNA eraldamine koosnes järgmistest etappidest:

1. Valkude hüdrolüüs: 1.5 ml tuubis segati kokku 500 µl süljeproovi ja 20 µl proteinaas K lahust (20 mg/ml) ning inkubeeriti segu termomikseris 48°C juures (*shaking*).

2. DNA seondumine filtrile: proovid segati 200 µl sidumispuhvriga (Binding Buffer A), asetati need filtrile, hoiti 1 min toatemperatuuril ja tsentrifuugiti 2 min 12 000 rpm juures.

3. Filtri pesemine: lisati järjestikku 500 µl esimest pesupuhvrit (Wash Buffer I) ja 600 µl teist pesupuhvrit (Wash Buffer II). Pärast iga puhvri lisamist tsentrifuugiti proove 2 min 12 000 rpm juures ja vahetati tuubi filtri all. Teiste puhvri lisamist ja sellele järgnevat tsentrifuugimist viidi läbi kaks korda. Seejärel tsentrifuugiti proove 5 min 13 400 rpm juures, et saada lahti pesupuhvrite jääkidest.

4. DNA elueerimine: filtrile asetati 100 µl eelsoojendatud (48°C) destilleeritud vett või elueerimispuhvrit (Elution Buffer D) ja tsentrifuugiti proove 2 min 12 000 rpm juures.

2.2.2.3 DNA kontsentratsiooni mõõtmine

DNA kontsentratsiooni ja proovide puhtust mõõdeti spektrofotomeetri abil (NanoDrop ND-1000). Spektrofotomeetri tööprintsip seisneb proovi läbiva valguse spektri mõõtmisel ja selle alusel lahuse koostise määramisel. Tänu erinevate ainete neeldumisspektrite erinevusele saab määrata nende sisaldust ja vahekorda lahuses. Edasiseks tööks peab DNA kontsentratsioon proovides olema suurem kui 20 ng/µl. Kui mõõdetud kontsentratsioon oli väiksem, proove kontsentreeriti. Hinnati ka proovide puhtust: 260/280 nm lainepikkuste suhe näitab valgusaastust ja 260/230 nm lainepikkuste suhe näitab saastust soolade ja solventidega. DNA proovi peetakse „puhtaks“, kui selle 260/280 nm suhe on ~1.8 ja 260/230 nm suhe on ~2.0-2.2 [9]. Kui need suhted erinesid oluliselt optimaalsetest (antud töös otsustati oluliseks erinevuteks peeta 0.5 ühikut), teostati proovide lisapuhastus.

2.2.2.4 Proovide kontsentreerimine

Et tõsta vajadusel DNA kontsentratsiooni, aurustati liigset vedelikku. Selleks kuivatati proove avatud kaantega -40°C juures. 10 minutiga aurustus umbes 10 µl vedelikku. Jooksvalt mõõdeti DNA kontsentratsiooni proovides ja kontsentreeriti proove kuni kontsentratsioon oli suurem kui 20 ng/µl.

2.2.2.5 DNA täiendav puhastamine

Kui DNA proovid olid saastatud valkude, soolade ja solventidega, võis see raskendada edasist analüüsi. Et sellest üle saada, teostati proovide täiendav puhastamine. Selleks kasutati järgmiseid vahendeid: 1.5 ml tuube ja QIAamp MinElute kolonne (Qiagen). Igale proovile lisati 10 µl AW1 puhvrit ja 250 µl AW2 puhvrit, proove segati vortex-segajal 10 s, asetatati proovid filtrile ja tsentrifuugiti 1 min 8000 rpm juures. Seejärel eemaldati tuubist vedelik ning tsentrifuugiti tühja tuubiga 3 min 13 400 rpm juures kogu vedeliku eemaldamiseks. Elueerimiseks vahetati filtri all tuubi kogumistuubi vastu, lisati kolonnile 100 µl destilleeritud vett, hoiti proove 5 min toatemperatuuril ja tsentrifuugiti 1 min 13 400 rpm juures.

Lõplikud kontsentratsioonid ja puhtuse näitajad on toodud Tabelis 1. Keskmiselt kõrgeimad näitused olid veega elueeritud süljeproovidel. Andmed teiste proovide kohta ei ole toodud kuna neid edasises töös ei kasutatud.

Tabel 1. DNA kontsentratsioonid ja saastuse näitajad uuritud 12 proovi jaoks. DNA oli eraldatud süljeproovidest ja elueeritud veega

Proovi number	DNA kontsentratsioon (ng/μl)	Proovi puhtus (260/280 nm; 260/230 nm)
1	69.7	1.8; 1.0
2	48.0	1.7; 0.3
3	247.1	1.1; 1.1
4	96.2	0.8; 0.7
5	71.6	1.5; 0.7
6	83.5	1.9; 1.8
7	34.8	1.8; 1.1
8	220.0	1.7; 1.8
9	149.5	1.9; 1.5
10	41.2	1.4; 0.7
11	23.7	2.0; 1.2
12	26.5	1.9; 1.4

Proovide genotüpiseerimiseks (mtDNA haplogrupi määramiseks) kasutati kaht erinevat meetodit: kogu mtDNA genoomi sekveneerimist ja kontrollregiooni sekveneerimist edasise kodeerivas alas asuvate informatiivsete regioonide sekveneerimisega. Kuna kogu mtDNA sekveneerimiseks on vaja kõrget DNA kontsentratsiooni ja kõrgeid puhtuse näitajaid, siis analüüsiti selle meetodiga proove 1, 3, 6 ja 8. Ülejäänud proovidest (2, 4, 5, 7, 9, 10, 11, 12) sekveneeriti lühemaid fragmente (HVS-regioone ja kodeeriva ala informatiivseid regioone).

Esialgu valiti kogu mtDNA genoomi sekveneerimiseks ka 9. proovi, mille puhtuse ja kontsentratsiooni näitajad olid üks parimaid (Tabel 1), kuid selle jaoks pikkade fragmentide amplifitseerimine ebaõnnestus (2.2.3.3), seetõttu analüüsiti seda samamoodi nagu madalama kvaliteediga proove.

2.2.3 Täisgenoomne mtDNA järjestuse määramine

Järjestuse määramine teostati Rieder'i meetodika järgi (Rieder jt., 1998). Meetod põhineb kogu mtDNAd katva nelja pika (u. 4000 ap) lõigu amplifitseerimisel ja sekveneerimisel; sekveneerimiseks jaotati iga lõik kuueks lühemaks kattuvaks lõiguks.

2.2.3.1 Pikkade fragmentide amplifitseerimine

Kogu mtDNA amplifitseeriti nelja praimeripaari abil (1F, 6R; 6F, 12R; 12F, 18R; 18F, 24R; Rieder jt., 1998). F tähistab siin ja edaspidi päripidist (*forward*) praimerit ja R äraspidist (*reverse*) praimerit. Kasutati DNA polümeraasi Phusion Hot Start II High Fidelity DNA Polymerase ja puhvrit 5X Phusion HF Buffer (Thermo Scientific). Kõikide töös teostatud PCR-reaktsioonide ja amplifitseerimisprodukti puhastamise jaoks kasutati Biometra UNO II Thermal

Cycler termostaate. MtDNA pikkade fragmentide amplifitseerimiseks kasutatud PCR-segu koostis ja termostaadi programm on toodud tabelis 2.

Tabel 2. Kogu mtDNA amplifitseerimiseks kasutatud PCR-segu (A) ja PCR-programm (B). F tähistab päripidist (*forward*) ja R äraspidist (*reverse*) praimerit

A. PCR-segu (kogumaht 25 µl)	µl	B. PCR-programm	kestvus
5X puhver (HF)	5	1.eelkuumutus: 98°C	30 s
DNA polümeraas (2 U/µl)	0.25	2.denaturatsioon: 98°C	15 s
dNTP Mix (20 µM)	0.25	3.praimerite seondumine: 56°C	30 s
vesi	14.5	4. DNA süntees: 72°C	6 min
praimerid (10 pmol/µl) (F + R)	2 + 2	35 tsüklit (etapid 2. – 4.)	
		5. järelsüntees: 72°C	7 min
DNA proov	1	4°C	∞

2.2.3.3 Amplifitseerimisprodukti visualiseerimine geelelektroforeesi abil

PCR-reaktsiooni produktide olemasolu ja hulga kontrollimiseks voolutati 2.5 µl PCR-produkti 1.5% agarosgeelil 0.5x TBE puhvril (54 mM Tris-Boraat, 0.05 M EDTA), geelile lisati 0.5 µg/ml etiidiumbromiidi (DNAGA interkaleeruv komponent, mis fluorestseerub UV-valguses) 130 V pingel juures ning seejärel visualiseeriti tulemust UV-kiirguses (Uvitec Gel Documentation System).

2.2.3.4 Amplifitseerimisprodukti puhastamine

Amplifitseerimisprodukti kõrval sisaldab PCR-segu desoksünukleotiide (dNTP) ja praimerite jääke, mis võivad vähendada järgneva sekveneerimise efektiivsust. Selle vältimiseks puhastati amplifitseerimisprodukti kasutades kahte ensüümi: kreveti aluselist fosfataasi (SAP ehk *Shrimp Alkaline Phosphatase*), mis lagundab seondumata dNTPsid, ja eksonukleaasi ExoI (*Thermo Scientific Exonuclease I* (20 U/L)), mis lagundab seondumata primereid (mõlema tootja on Thermo Fisher Scientific). Igasse proovi lisati 1 µl ExoSap segu, mis sisaldab 9 osa SAP ja 1 osa ExoI ning kasutati järgmist termostaadi programmi:

1. ensüümide toimimine: 37°C – 15 min
2. ensüümide inaktivatsioon: 80°C – 20 min

2.2.3.5 Sekveneerimine

Sekveneerimine teostati Sanger'i meetodil (Sanger jt., 1977) kasutades BigDye Terminator v3.1 Cycle Sequencing Kit'i (ThermoFisher Scientific). PCR-reaktsiooni ja produkti puhastamise viis läbi käesoleva töö autor ja sekveneerimisprodukti kapillaarset geelelektroforeesi teostati Eesti Biokeskuse (EBK) DNA sekveneerimise tuumiklaboris

masinal ABI3700 (Applied biosystems). Täissekveneerimisel kasutati iga 4000 ap pikka PCR-produkti jaoks 6 paari kattuvaid primereid: näiteks kui amplifitseerimiseks oli kasutatud paari 1F/6R, siis sekveneerimiseks kasutati 1F, 1R; 2F, 2R; 3F, 3R; 4F, 4R; 5F, 5R; 6F, 6R. MtDNA pikkade fragmentide sekveneerimiseks kasutatud PCR-segu koostis ja termostaadi programm on toodud Tabelis 3.

Tabel 3. Kogu mtDNA sekveneerimiseks kasutatud PCR-segu (A) ja PCR-programm (B)

A. PCR-segu (kogumaht 10 µl)	µl	B. PCR-prorgamm	kestvus
ensüüm (BigDye Terminator v3.1 Ready Reaction Mix)	1	1. eelkuumutus: 95°C	1 min
		2. denaturatsioon: 95°C	15 s
Big Dye puhver (5x)	2	3. praimerite seondumine: 52°C	10 s
vesi	4	4. DNA süntees: 60°C	1 min
praimer (1 pmol/µl)	2	30 tsükli (etapid 2. – 4.)	
PCR-produkt	1	4°C	∞

2.2.3.6 Sekveneerimiseks kasutatava PCR-produkti puhastamine

Sekveneerimiseks amplifitseeritud DNA kõrval sisaldab PCR-segu teisi aineid (nt. primereid ja soolasisi), millest tuleb produkti puhastada enne kapillaarse geelelektroforeesi läbiviimist. Produkti puhastamine sisaldas järgmisi etappe:

1. DNA sadestamine: igale proovile lisati 2 µl ammooniumatsetaat-dekstraan segu (1 osa 1.5 M CH₃COONH₄ [pH>8]/EDTA (250 mM) ja 1 osa 20 mg/ml dekstraani) ja 30 µl 96% EtOH ning inkubeeriti 20 min -20°C juures. Ammooniumatsetaadi toimel muutub DNA lahustumatuks soolaks ja roosa dekstraan aitab sadet visualiseerida.

2. DNA puhastamine: proove tsentrifuugiti 15 min 13 000 rpm juures ja eemaldati supernatandi veejoapumba abil. Seejärel kanti sademele 200 µl külma (-20°C) 70% EtOH, tsentrifuugiti 5 min 13 000 rpm juures ja eemaldati supernatanti. Korrati viimast etappi. Proovid jäeti u. 10 minutiks 37°C termostaati avatud kaantega kuivama, et piirituse jäägid aurustuksid.

3. DNA lahustamine: kuivale sademele kanti 10 µl formamiidi lahust (MegaBACE loading solution: 70% formamiid/ 1 mM EDTA segu) ja lasti DNAl 5 min toatemperatuuril lahustuda. Lõpetuseks kanti proovid plaadile ja viidi EBK DNA sekveneerimise tuumiklaborisse, kus teostati kapillaarset geelelektroforeesi.

2.2.4 Lühikeste fragmentide amplifitseerimine ja sekveneerimine

HVS-regioonide ja kodeeriva piirkonna informatiivsete regioonide järjestuse määramine teostati amplifitseerimise (Tabel 4) ja edasise sekveneerimise teel. MtDNA kontrollregioonis asuvatele hüpervarieeruvatele piirkondadele (HVSII np vahemikus 57 – 372 ja HVSI np vahemikus 16 024 – 16 383 [1]) on iseloomulikud sagedased nukleotiidide asendused. Need on

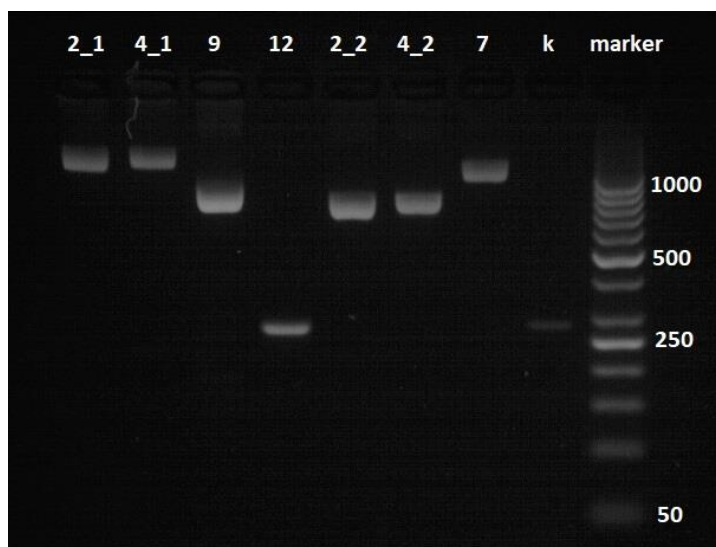
suhteliselt lühikesed fragmendid (315 ja 359 ap) ja seepärast õnnestub nende sekveneerimine ka madalama kvaliteediga proovidest. Juhul kui haplogruppi ei olnud võimalik määrata piisava täpsusega HVS-regioonides esinevate asenduste alusel, sekveneeriti täiendavalt kodeeriva ala piirkondi, kus asusid haplogruppide määramiseks olulised positsioonid. Amplifitseerimiseks kasutati 5x FIREPol Master Mix Ready to Load segu 12.5 mM MgCl₂ sisaldusega (Solis BioDyne).

Tabel 4. MtDNA lühikeste fragmentide amplifitseerimiseks kasutatud PCR-segu (A) ja PCR-programm (B). F tähistab päripidist (*forward*) ja R äraspidist (*reverse*) praimerit

A. PCR-segu (kogumaht 15 µl)	µl	B. PCR-prorgamm	kestvus
5X Master Mix	3	1. eelkuumutus: 95°C	3 min
vesi	9.4	2. denaturatsioon: 95°C	30 s
praimerid (10 pmol/µl) (F + R)	0.3 + 0.3	3. praimerite seondumine: 53°C (või 56°C)*	30 s
proov DNAGA	2	4. DNA süntees: 72°C	40 s
		35 tsüklit (etapid 2. – 4.)	
		5. järelsüntees: 72 °C	3 min
		4°C	∞

*Märkus: Rieder'i ja kaasautorite (Rieder jt., 1998) disainitud praimerite seondumine toimub 56 °C juures ja muude kasutatud praimerite seondumine toimub 53 °C juures

Amplifitseerimisprodukti visualiseerimine teostati samamoodi nagu on kirjeldatud punktis 2.2.3.3 järgmiste muudatustega: kasutati 3% agarosgeeli ja pinget 100 V. Näide agarosgeelil visualiseeritud amplifitseeritud fragmentidest on toodud joonisel 6.



Joonis 6. PCR-produktide visualiseerimine 3% agarosgeelil. Amplifitseerimiseks kasutati järgmiseid praimereid: 7793F, 9151R (2_1 ja 4_1 jaoks); 7458F, 8366R (9 jaoks); 12541F, 12818R (12 jaoks); 13F, 13R (2_2 ja 4_2 jaoks); 22F, 22R (7 jaoks). k – negatiivne kontroll, kuhu DNAd ei lisatud. Marker – Thermo Scientific GeneRuler 50 ap DNA Ladder.

HVS-regioonide ja kodeerivas alas asuvate regioonide amplifitseerimisproduktide puhastus, sekveneerimine ja sellele eelnev sekveneerimiseks kasutatud PCR-produkti puhastus teostati samamoodi nagu pikkade fragmentide jaoks (2.2.3.4 – 2.2.3.6). Sekveneerimiseks kasutati samu praimerid mida kasutati vastavate fragmentide amplifitseerimiseks. Tabelis 5 on toodud HVS-regioonide koordinaadid mtDNA1, nende amplifitseerimiseks ja sekveneerimiseks kasutatud praimerite alguspositsioonid ja järjestused.

Tabel 5. HVS-regioonide paiknemine mtDNA genoomis [1] ja nende järjestuse määramiseks kasutatud praimerid. F tähistab päripidist (*forward*) ja R äraspidist (*reverse*) praimerit

Ala	F praimer	R praimer	F primeri järjestus	R primeri järjestus
HVSI (16024 – 16383)	15795F	16498R	5' CTC CAC CAT TAG CAC CCA AAG 3'	5' CCT GAA GTA GGA ACC AGA TG 3'
HVSII (57 – 372)	16453F	727R	5' CCG GGC CCA TAA CAC TTG GG 3'	5' AGG GTG AAC TCA CTG GAA CG 3'

Tabelis 6 on näidatud proovid, mille jaoks oli haplogrupi täpsemaks määramiseks vajalik peale kontrollregiooni ka kodeeriva regiooni informatiivsete positsioonide sekveneerimine.

Tabel 6. Kodeeriva regiooni uuritud alad, nende järjestuse määramiseks kasutatud praimerid ja nendes asuvate informatiivsete positsioonide põhjal kontrollitavad haplogrupid. F tähistab päripidist (*forward*) ja R äraspidist (*reverse*) praimerit; Ta (*annealing temperature*) – praimerite seondumistemperatuur; inf. pos. – informatiivne positsioon; Hg – haplogrupp

Proo-vid	F primeri järjestus	R primeri järjestus	Ta, °C	Hõlmatud regioon	Inf. pos.	Hg
2, 4	5' GCC ATC ATC CTA GTC CTC ATC 3'	5' TTA AGG CGA CAG CGA 3'	53	7793 - 9151	8271, 8537	H1a2, H1a8
2, 4	13F: 5' TTT CCC CCT CTA TTG ATC CC3'	13R: 5' GTG GCC TTG GTA TGT GCT TT 3'	56	8621 - 9397	9341, 8839	H1a4, H1a5
7	22F: 5' TGA AAC TTC GGC TCA CTC CT 3'	22R: 5' AGC TTT GGG TGC TAA TGG TG 3'	56	14856 - 15978	15454	HV0e
9	5' GAA TCG AAC CCC CAA AGC TGG TTT CAA GC 3'	5' TTT CAC TGT AAA GAG GTG TTGG 3'	53	7458 - 8366	7864	W1
9	11F: 5' ACG CCA AAA TCC ATT TCA CT 3'	11R: 5' CGG GAA TTG CAT CTG TTT TT 3'	56	7148 - 8095	7864	W1
11, 12	5' GCC ACA ACC CAA ACA 3'	5' CGG GCG TAT CAT CAA 3'	53	12541 - 12818	12705	R0
11, 12	10F: 5' CTC TTC GTC TGA TCC GTC CT 3'	10R: 5' AGC GAA GGC TTC TCA AAT CA 3'	56	6469 - 7315	7094, 7028	HV4, H

2.2.5 Andmete analüüs

Sekveneerimise tulemusel saadud mtDNA järjestused liideti iga uuritud indiviidi jaoks kokku arvutiprogrammi ChromasPro [10] abil, lineariseerides neid RSRS (*Reconstructed Sapiens Reference Sequence*) referentsjärjestusele (Behar jt., 2012). Järgnevalt tuvastati referentsjärjestusest erinevad positsioonid ehk nukleotiidsed asendused ning määrati nende alusel haplogrupp. Selleks kasutati inimese mtDNA nomenklatuuri Phylotree (van Oven ja Klayser, 2009) versiooni mtDNA tree Build 17 (18 Feb 2016)).

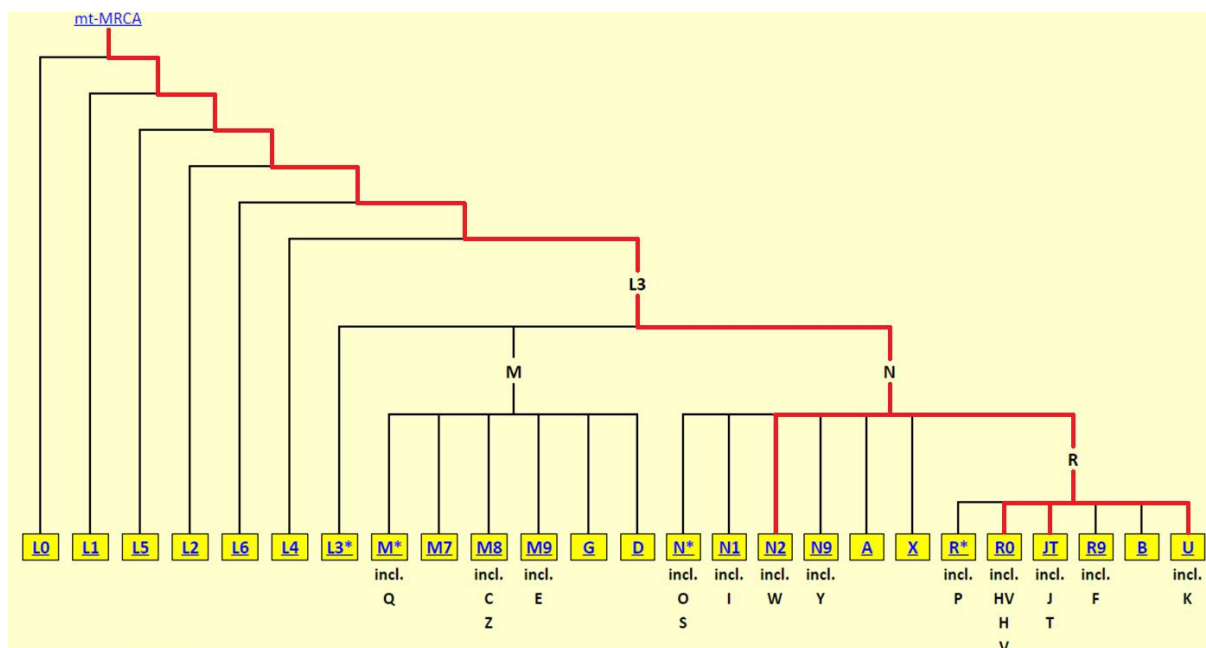
Pärast haplogruppide määramist kõrvutati neid olemasoleva andmestikuga eestlaste populatsioonis esinevate mtDNA haplogruppide kohta. Viimaseks andmeanalüüsi etapiks sai arheoloogide mtDNA järjestuste võrdlemine vana DNAGA, mille saasteallikaks nad võisid osutada. Andmeanalüüsi täpsem kirjeldus on esitatud punktides 2.3.1 – 2.3.3.

2.3 Tulemused ja arutelu

Töös uuritud 12 DNA proovi pärinesid 11 indiviidilt (10 arheoloogilt ja töö autorilt kahes korduses). Kahest eraldamismeetodist osutus paremaks s.t. andis parema kontsentratsiooni ja puhtusega proove süljest eraldamine ja nendest proovidest olid omakorda paremad näitajad veega elueeritud DNA proovidel. Proovides sisalduv mtDNA amplifitseeriti ja sekveneeriti. Saadud järjestuste põhjal määrati mtDNA haplogrupp iga proovi jaoks. Sellele järgnes edasine fülogeneetiline ja võrdlusanalüüs.

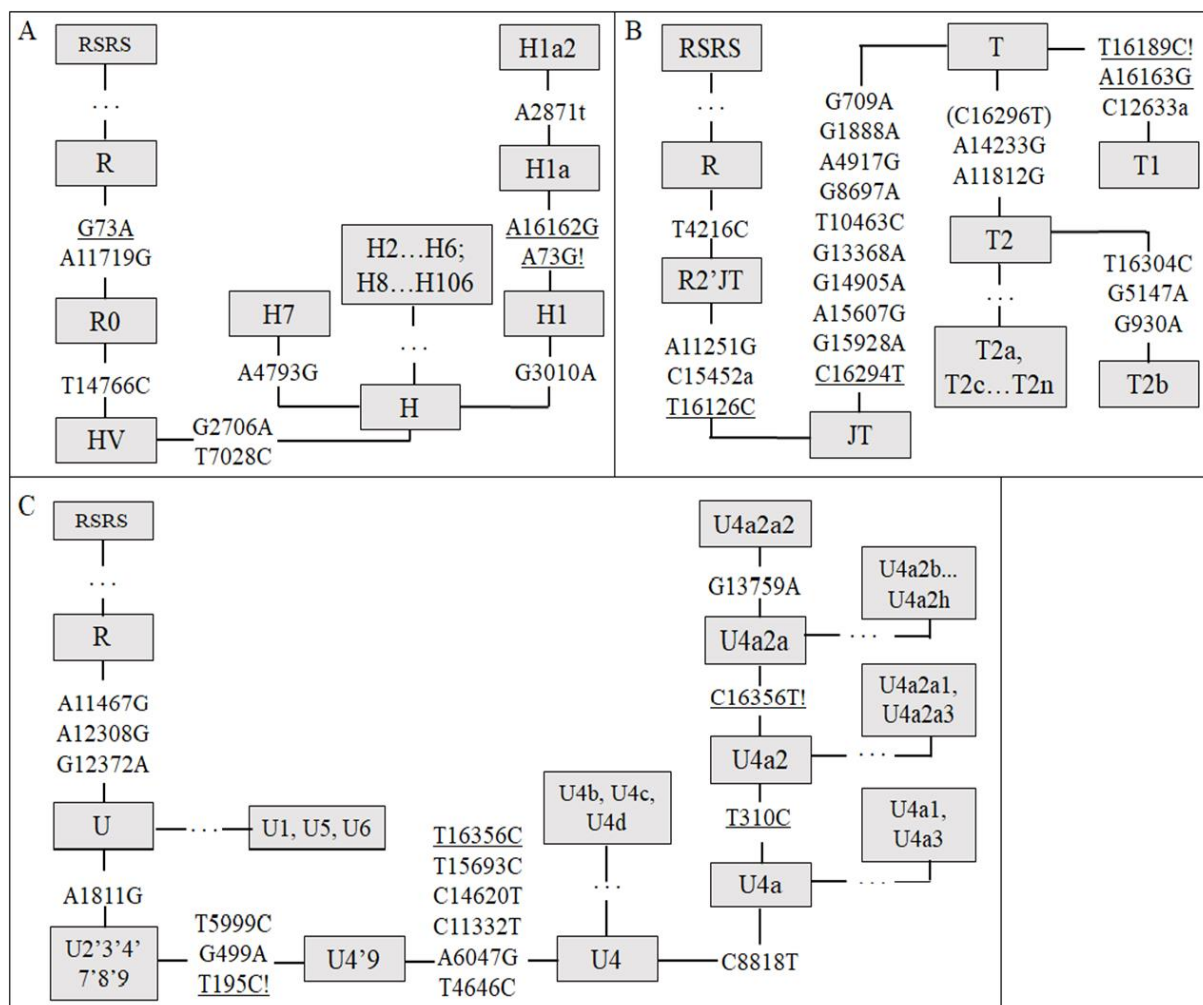
2.3.1 Arheoloogide mtDNA haplogrupid

MtDNA haplogrupid määrati kõikide uuritud proovide jaoks (vt. Lisa 1). Analüüsi tulemustest selgus, et viie uuritud indiviidi mtDNA haplogrupiks on H (proovid 1, 2, 3, 4, 11/12), kahe – T (8, 10), ühe – J (5), ühe – U (6), ühe – HV0 (7) ja ühe – W1 (9). Joonisel 7 on skemaatiliselt esitatud uuritud proovide haplogruppide paiknemine mtDNA fülogeneetilisel puul. Nagu näha, kuuluvad kõik järjestused suurde N haplogruppi ning kõik peale ühe kuuluvad selle alamharu R koosseisu.



Joonis 7. 12 uuritud proovi mtDNA järjestuste paiknemine mtDNA fülogeneesipuul (van Oven ja Klayser, 2009). Punase värviga on märgistatud vastavad harud ja nende fülogeneetiline asend puu juure, s. t. viimase ühise mitokondriaalse esivanema (*mitochondrial most recent common ancestor*, mt-MRCA) suhtes

Proovide haplogruppide määramiseks kasutati nende mtDNA järjestustes leiduvaid nukleotiidseid asendusi, mida võrreldi Phylotree andmebaasis (van Oven ja Kayser, 2009) olevatega. Allpool on näidatud detailsed fülogeneesipuud (joonis 8) iga leitud haplogrupi kaupa täielikult sekveneeritud proovide jaoks (1, 3, 6, 8).



Joonis 8. A. Esimese (H7) ja kolmanda (H1a2) proovi haplogruppide paiknemine haplogrupi H fülogeneetilisel puul. B. Kaheksanda proovi haplogrupi (T2b) paiknemine haplogrupi T fülogeneetilisel puul. C. Kuuenda proovi haplogrupi (U4a2a2) paiknemine haplogrupi U fülogeneetilisel puul. Transitsioonid on tähistatud kahe suure tähega positsiooni numbriga nende vahel (nt. T7028C), transversioonid – suure ja väikese tähega (nt. A2871t), tagasiasendused on tähistatud hüüumärgiga (nt. A73G!), asendused kontrollregioonis on alla joonitud (nt. G73a). Punktir (...) tähistab nukleotiidseid asendusi nende rühmade jaoks, millesse proov ei kuulu ja mida ei ole seetõttu täpsemalt iseloomustatud. Asendused alates puu juurest kuni R haplogrupini on kõikide proovide jaoks ühesugused, seetõttu ei ole nad ruumi kokkuhoiu mõttes välja kirjutatud

2.3.2 Arheoloogide mtDNA võrdlus Eesti kaasaegse elanikkonna andmestikuga

Arheoloogide mtDNA haplogruppe kõrvutati olemasoleva informatsiooniga eestlaste mtDNA mitmekesisuse kohta. Võrdluseks kasutati kaht allikat: 2004. aasta uuringut (Pliss jt., 2004) ja EBK mtDNA andmebaasi. Pliss ja kaasautorid näitasid oma uuringus, et haplogrupp H esineb 43.5% eestlastest, U – 26.9%, järgmised levinuimad on J (10.3%) ja T (7.8%) ning üsna tihti esinevad eestlastel ka V (3.9%), W (2.4%) ja HV (1.5%). EBK mtDNA andmebaasis leiduv informatsioon oli samuti nende andmetega kooskõlas.

Kuigi antud töös analüüsiti vaid 11 erinevat proovi, mistõttu on valim liiga väike ega võimalda esinemissageduste võrdlemist, võib täheldada, et selles on esindatud paljud Eesti sagedaseimad haplogrupid ja suurim arv proove on määratud kuuluvaks haplogruppi H, mis esineb eestlastel (ja eurooplastel üldiselt) kõrgeima sagedusega (Pliss jt., 2004).

Seega, võrreldes arheoloogide mtDNA haplogruppe mtDNA mitmekesisusega kaasaegsel Eesti elanikkonnal jõuti järeldusele, et uuritud proovide haplogruppide kogum on olemasoleva andmestikuga kooskõlas.

2.3.3 Arheoloogide mtDNA võrdlus Eesti keskaegse elanikkonna andmestikuga

Peale haplogruppide määramist teostati arheoloogide 11 mtDNA haplogrupi võrdlus 35 haplogrupiga, kuhu kuulusid arheoloogide väljakaevatud luumaterjalist eraldatud vana DNA järjestused (luumaterjal pärineb Eesti territooriumi keskaegse (13.-15. saj.) valimi hulgast; EBK vana DNA töögrupi avaldamata andmed). Võrdlusanalüüsi tulemusel tuvastati mõningad kokkulangevused arheoloogide ja iidsete proovide haplogruppide vahel (Lisa 2).

Nelja vana DNA haplogrupp (T2b) ja arheoloogide 8. proov langevad samasse haplogruppi, kuid hulk individuaalseid asendusi osutab sellele, et proovid kuuluvad erinevatele indiviididele. Ühe vana DNA järjestuse haplogrupiks määrati U4, mille alamharuks on arheoloogide 6. proovi haplogrupp (U4a2a2), kuid võrreldes nukleotiidseid asendusi on näha, et proovid on esindatud erinevate haplotüüpidega. Veel üks iidne proov kuulub haplogruppi J1b1a3, mille sõsarhaplogruppi J1b1a1 kuulub 5. kaasaegse proovi mtDNA. Nende kahe järjestuse vahel on samuti palju erinevusi. Ülejäänud vana DNA proovid ei esine haplogruppide kokkulangevusi arheoloogide mtDNA proovidega. Võrdlusanalüüsi põhjal tehti järelduse, et käesolevas uuringus saadud kaasaegse mtDNA järjestused ei ole käsitletud vana DNA saastajaks.

Võrdlusanalüüs luumaterjaliga töötavate inimeste DNAGA, mis viidi läbi käesolevas töös, on vaid osa sellest luumaterjalist eraldatud mtDNA komplekssest autentsuse analüüsist. Peale vana DNA järjestuste võrdlemise arheoloogide mtDNAGA teostati EBK vana DNA uurimisrühma teadlaste poolt ka teisi autentsuse kontrollimiseks vajalikke teste. Esiteks teostati vana DNA saastustest (kirjeldatud punktis 1.4), mis näitas, et vana DNA proovides on saastust alla 3%. Teiseks teostati vana DNA kahjustuste analüüs, mis näitas, et vana DNA proovide kahjustuste muster (*damage pattern*) vastab oodatule. Seega, käesoleva uuringu ja täiendtestide tulemustest võib järeldada, et antud vana DNA proovides saastus kaasaegse DNAGA kas puudub või esineb koguses, mis ei sega vana DNA autentsete järjestuste saamist.

Kokkuvõte

Vanast bioloogilisest materjalist eraldatud DNA autentsuse tõestamine on inimese vana DNA uuringute üks olulisemaid osi. Üks etappidest, kus kaasaegse inimese DNA võib vanu proove saastada, on iidse bioloogilise materjali väljakaevamine ja ettevalmistamine molekulaargeneetiliseks analüüsiks – saasteallikaks võivad osutuda arheoloogid. Käesoleva töö peamiseks eesmärgiks oli Eestis arheoloogiliste väljakaevamiste käigus saadud luumaterjalist eraldatud vana DNA puhtuse kontrollimine. Selleks teostati eesti arheoloogide rühma liikmete mtDNA eraldamine ja molekulaargeneetiline analüüs, misjärel võrreldi seda keskaegsest luumaterjalist eraldatud mtDNAGA.

Nukleotiidsete asenduste analüüsi tulemusel määrati mtDNA haplogrupid kõikide proovide jaoks. Arheoloogide haplogruppide kõrvutamine olemasoleva andmestikuga kaasaegse eestlaste populatsiooni kohta näitas, et saadud tulemused on kooskõlas üldise mtDNA mitmekesisusega Eesti elanikel. Vaatamata sellele, et mõnede vanade ja kaasaegsete proovide mtDNAd kuulusid samasse haplogruppi, erinesid nende proovide haplotüübid olulisel määral ja täielikke kokkulangevusi ei esinenud.

Käesoleva töö tulemused võimaldavad väita, et vana DNA proovide saastust arheoloogide geneetilise materjaliga ei ole tõenäoliselt toimunud. Samas on proovidega töötavad inimesed vaid üks paljudest võimalikest saasteallikatest ja täielik veendumine vana DNA järjestuste autentsuses vajab kompleksset lähenemist: antud töös käsitlevate proovide jaoks olid varasemalt teostatud veel saastustest ja kahjustusmusteri analüüs, mis samuti kinnitasid proovide puhtust.

Suurema osa DNA proovide mtDNA haplogrupi määramiseks kasutati informatsiooni nii mtDNA kontrollregioonist kui ka kodeeriva regiooni informatiivsetest aladest. Kui aga võrreldi arheoloogide mtDNA järjestusi vana DNAGA, piisas enamasti kontrollregioonis olevatest erinevustest väitmaks, et proovid kuuluvad erinevatesse haplogruppidesse, s.t. saastuse välistamiseks. Samas peab mainima, et mtDNA kogu genoomi sekveneerimine annab palju rohkem informatsiooni, kui ainult kontrollregiooni analüüs, võimaldades järjestuse täpsemat klassifitseerimist mtDNA fülogeneetilisel puul.

Kõik töös püstitatud eesmärgid saavutati; saadud tulemused täiendavad Eesti keskaegse elanikkonna geneetilise mitmekesisuse uurimiseks kavandatud projekti raames tehtud vana DNA autentsuse määramise teste.

Detection of aDNA contamination through comparative analysis of modern and ancient mitogenomes

Maria Maloverjan

Summary

Ancient DNA (aDNA) is genetic material extracted from tissues of dead organisms, and human aDNA plays an important role in evolutionary and population studies of human species. Because of its poor chemical stability and, consequently, low quality and quantity in old remains, aDNA is much more difficult to analyze than modern DNA, and that is why ancient samples are subjected to contamination. During history of aDNA studies, contamination with modern genetic material has led to a number of articles based on false results, so nowadays there are strict rules and precautionary measures to avoid contamination, and, in addition, all aDNA sequences got must be verified.

The aim of the present work was to carry out a comparative analysis of modern and ancient mtDNA sequences as part of the complex procedure of aDNA contamination detection. For this purpose, DNA was extracted from biological samples of archaeologists, who have provided bone material for further aDNA analysis. Archaeologists' mtDNA was amplified and sequenced. Then, phylogenetic analysis was performed and haplogroups for all samples were determined. Finally, comparative analysis with aDNA mitogenomes of the samples of archaeologists was carried out. In addition, the mtDNA's sequences of archaeologists were compared with general mtDNA diversity among Estonians to see if the results of the present work correspond to the existing data.

It was shown that haplogroups of the studied mtDNAs of archaeologists are common in modern Estonians. During comparative analysis of modern and ancient mitogenomes, it could be observed that some haplogroups were shared between archaeologists and ancient samples. However, there was always a number of nucleotide substitution differences between modern and ancient sequences, so no complete matches of haplotypes were detected. Thus, according to our results, no archaeologists' mtDNA was found in ancient samples hence mtDNA-based contamination test was negative for studied samples.

Although mtDNA comparative analysis can be considered sufficient to exclude contamination of aDNA with genetic material of people who worked with bone samples, it is only a part of a complex aDNA authentication procedure. For aDNA sequences studied, additional tests were performed, too, and the purity of samples was confirmed. Therefore, it was concluded that no contamination of ancient samples occurred and aDNA can be considered authentic and suitable for further population-genetic and evolutionary analyses.

Kasutatud kirjandus

- Giles, R. E., Blanc, H., Cann, H. M. and Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 77(11):6715-9.
- Brown, W. M., George, M. Jr. and Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 76(4):1967-1971. Genetics.
- Jodre, L. B., Bramshad, M. and Rogers, A. R. (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 20(2):126-36.
- Anderson, S., Bankier, A. T., Barrell B. G., ... Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.
- Andrews, R. M., Kubacka, I., Chinnery, P.F., Lightowlers, R. N., Turnbull, D. M. and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.
- Birky C. W. Jr. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet* 35:125-48.
- Pesole, G., Gissi, C, De Chirico, A. and Saccone, C. (1999). *J Mol Evol* 48(4):427-34.
- Rooney, J. P., Ryde, I. T., Sanders, L. H., Howlett, E. H., Colton, M. D., Germ, K. E., Mayer, G. D., Greenamyre, J. T. and Meyer J. N. (2015). PCR Based Determination of Mitochondrial DNA Copy Number in Multiple Species. *Methods Mol Biol* 1241: 23–38.
- Bandelt, H. J., Forster, P., Sykes, B. C. and Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141(2):743-53.
- Brandstätter, A., Salas, A., Niederstätter, H., Gassner, C., Carracedo, A., Parson, W. (2006). Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* 27(13):2541-50.
- Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M. and Pääbo, S. (1997). Neandertal DNA Sequences and the Origin of Modern Humans. *Cell* 90:19–30.
- Adachi, N., Kakuda, T., Takahashi, R., Kanzawa-Kiriyama, H. and Shinoda, K. (2017). Ethnic derivation of the Ainu inferred from ancient mitochondrial DNA data. *Am J Phys Anthropol* 165:139-148.
- Curat, M. and Excoffier, L. (2004). Modern Humans Did Not Admix with Neanderthals during Their Range Expansion into Europe. *PLoS Biol* 2(12): e421.
- Higuchi, R., Bowman, B., Freiburger, M., Ryder, O. A. and Allan C. Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282–284.

- Ingman, M., Kaessmann, H., Pääbo, S. and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708-13.
- van Oven, M. and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):E386-E394. <http://www.phylotree.org>
- Underhill, P. A. and Kivisild, T. (2007). Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. *Annu Rev Genet* 41:539–64.
- Brown, T. A. (2002). *Genomes*, 2nd edition. Chapter 1.2 The Human Genome. ISBN-10: 0-471-25046-5.
- Quintana-Murci, L. and Fellous, M. (2001). The human Y chromosome: the biological role of a “functional wasteland”. *J Biomed Biotechnol* 1(1):18–24.
- Wright, S. (1951). The genetical structure of populations. *Ann Eugen* 15(4):323-54.
- Weir, B. S., Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358-1370.
- Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical magazine, Series 6*, 2(11):559-572.
- Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. *J Educ Psychol* 24(6,7):417-441, 498-520.
- Alexander, D. H., Novembre, J. and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655-64.
- Crawford, D. C. and Nickerson, D. A. (2005). Definition and clinical importance of haplotypes. *Annu Rev Med* 56:303–20.
- Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D. and Myers, S. (2014). A genetic atlas of human admixture history. *Science* 343(6172):747–751.
- Browning S. R. (2008). Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum Genet* 124(5):439-50.
- Novembre, J., Johnson, T., Bryc, K., ... Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature* 456:98–101.
- Pagani, L., Lawson, D. J., Jagoda, E., ... Metspalu, M. (2015). *Nature* 538:238–242.
- Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E. and Allentoft, M. E. (2017). Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PLoS ONE* 12(1):e0170940.
- Pääbo, S., Gifford, J. A. and Wilson A. C. (1988). Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Res* 16(20):9775-87.
- Allentoft, M. E., Sikora, M., Sjögren, K.-G., ... Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.

- Haak, W., Lazaridis, I., Patterson, N., ... Reich, D. (2015). Massive migration from the steppe is a source for Indo-European languages in Europe. *Nature* 522:207–211.
- Mathieson, I., Lazaridis, I., Rohland, R., ... Reich, D. (2016). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528:499–503.
- Meyer, M., Arsuaga, J. L., de Filippo, C., ... Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531(7595):504-7.
- Orlando, L., Gilbert, M. T. and Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. *Nat Rev Genet* 16(7):395-408.
- Reich, D., Green, R. E., Kircher, M., ... Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.
- Meyer, M., Kircher, M., Gansauge, M. T., ... Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222-6.
- Green, R. E., Krause, J., Briggs, A. W., ... Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science* 328(5979):710-722.
- Prüfer, K., Racimo, F., Patterson, N., ... Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43-9.
- Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., ... Reich, D. (2018). The genomic history of southeastern Europe. *Nature* 555:197–203.
- Lazaridis, I., Patterson, N., Mittnik, A., ... Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409-13.
- Bramanti, B., Thomas, M. G., Haak, W., ... Burger, J. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326(5949):137-40.
- Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M. T. P., Götherström, A. and Jakobsson, M. (2012). Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* 336(6080):466-469.
- Thomas, W. K. and Pääbo, S. (1993). DNA sequences from old tissue remains. *Methods Enzymol* 224:406-419.
- Pinhasi, R., Fernandes, D., Sirak, K., ... Hofreiter, M. (2015). Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS ONE* 10(6):e0129102.
- Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E. and Allentoft, M. E. (2017). Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PLoS ONE* 12(1):e0170940.
- Margaryan, A., Hansen, H. B., Rasmussen, S., ... Allentoft, M. E. (2018). Ancient pathogen DNA in human teeth and petrous bones. *Ecology and Evolution* 8(6):3534-3542.

- Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., Evershed, R. P., Possnert, G. and Pääbo, S. (1998). Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* 281(5375):402-6.
- Rasmussen, M., Guo, X., Wang, Y., ... Willerslev, E. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334(6052):94-8.
- Gilbert, M. T., Kivisild, T., Grønnow, B., ... Willerslev, E. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320(5884):1787-9.
- Rasmussen, M., Li, Y., Lindgreen, S., ... Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.
- Gilbert, M.T., Wilson, A. S., Bfonce, M., ... Cooper, A. (2004). Ancient mitochondrial DNA from hair. *Curr Biol* 14(12):R463-4.
- Gilbert, M. T., Barnes, I., Collins, M. J., Smith, C., Eklund, J., Goudsmit, J., Poinar, H. and Cooper, A. Long-term survival of ancient DNA in Egypt: response to Zink and Nerlich (2003). *Am J Phys Anthropol* 128(1):110-4; discussion 115-8.
- Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314:644–645.
- Schuenemann, V. J., Peltzer, A., Welte, B., ... Krause, J. (2016). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nature Communications* 8, Article number: 15694 (2017).
- Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci USA* 86(6):1939-43.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*. 362(6422):709-15.
- Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. and Pääbo, S. (2001). Ancient DNA. *Nature Reviews Genetics* 2:353–359.
- Allentoft, M. E., Collins, M., Harker, D., ... Bunce, M. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc R Soc B*, published online, doi: 10.1098/rspb.2012.1745
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. and Pääbo, S. (2012). Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLoS ONE* 7(3):e34131.
- Burger, J., Hummel, S., Herrmann, B. and Henke, W. (1999). DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* 20:1722-1728.
- Nicholls, H. (2005) Ancient DNA Comes of Age. *PLoS Biol* 3(2):e56.

- Hansen, A. J., Mitchell, D. L., Wiuf, C., Paniker, L., Brand, T. B., Binladen, J., Gilichinsky, D. A., Rønn, R. and Willerslev, E. (2006). Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* 173(2):1175-9.
- Orlando, L., Ginolhac, A., Zhang, G., ... Willerslev, E. (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499:74–78.
- Willerslev, E. and Cooper, A. (2005). Ancient DNA. *Proc Biol Sci* 272(1558):3–16.
- Arsuaga, J. L., Martínez, I., Arnold, L. J., ... Carbonell, E. (2014). Neandertal roots: Cranial and chronological evidence from Sima de los Huesos. *Science* 344(6190):1358-63.
- Rohland, N. and Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols* 2:1756–1762.
- Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. and Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645-79.
- Dabney, J., Knapp, M. and Glocke, I. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *PNAS* 110(39):15758-15763.
- Willerslev, E. and Cooper, A. (2005). Ancient DNA. *Proc Biol Sci* 272(1558):3-16.
- Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J. and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *PNAS* 111(6):2229-2234.
- Bainbridge, M. N., Wang, M., Burgess, D. L., ... Gibbs, R. A. (2010). Whole exome capture in solution with 3 Gbp of data. *Genome Biology* R62.
- Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., ... Bustamante, C. D. (2013). Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am J Hum Genet* 93(5):852–864.
- Green, R. E., Krause, J., Ptak, S. E., ... Pääbo, S. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature* 444:330–336.
- García-Garcera, M., Gigli, E., Sanchez-Quinto, F., Ramirez, O., Calafell, F., Civit, S. and Lalueza-Fox, C. (2011). Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. *PLoS ONE* 2011;6(8):e24161.
- Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelsob, J. and Pääbo, S. (2012). DNA analysis of an early modern human from Tianyuan Cave, China. *PNAS* 110(6):2223-2227.

- Poinar, H. N., Schwarz, C., Qi, J., ... Schuster, S. C. (2006). Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science* 311(5759):392-4.
- Sampietro, M. L., Gilbert, M. T., Lao, O., Caramelli, D., Lari, M., Bertranpetit, J. and Lalueza-Fox, C. (2006). Tracking down human contamination in ancient human teeth. *Mol Biol Evol* 23(9):1801-7.
- Gilbert, M. T. P., Hansen, A. J., Willerslev, E., Turner-Walker, G. and Collins, M. (2006). Insights into the processes behind the contamination of degraded human teeth and bone samples with exogenous sources of DNA. *Int J Osteoarchaeol* 16:156–164.
- Cooper, A., Poinar, H. N. (2000). Ancient DNA: do it right or not at all. *Science* 289(5482):1139.
- Knapp, M., Clarke, A. C., Horsburgh, K. A. and Matisoo-Smith, E. A. (2012). Setting the stage – Building and working in an ancient DNA laboratory. *Annals of Anatomy* 194:3–6
- Morin, P. A., Chambers, K. E., Boesch, C. and Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Mol Ecol* 10(7):1835-44.
- Fulton, T. L. (2011) Setting Up an Ancient DNA Laboratory. *Methods in Molecular Biology* 840:1-11.
- Kemp, B. M. and Smith, D. G. (2005). Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci Int* 154(1):53-61.
- Brotherton, P., Endicott, P., Sanchez, J. J., Beaumont, M., Barnett, R., Austin, J. and Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35(17):5717-28.
- Caramelli, D., Milani, L., Vai, S., ... Barbujani, G. (2008). A 28,000 Years Old Cro-Magnon mtDNA Sequence Differs from All Potentially Contaminating Modern Sequences. *PLoS ONE* 3(7):e2700.
- Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J. and Mattias Jakobsson. (2013). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *PNAS* 111(6):2229–2234.
- Ermini, L., Olivieri, C., Rizzi, E., ... Rollo, F. (2008). Complete Mitochondrial Genome Sequence of the Tyrolean Iceman. *Current Biology* 18(21):1687-1693.
- Woodward, S. R., Weyand, N. J. and Bunnell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science* 266(5188):1229-32.

- Vreeland, R. H., Rosenzweig, W. D. and Powers, D. W. (2000). Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* 407(6806):897-900.
- Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D. and Zurawski, G. (1990). Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature* 344:656–658.
- Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A. and Poinar Jr, G. O. (1992). Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature* 363:536–538.
- Walden, K. K. and Robertson, H.M. (1997). Ancient DNA from Amber Fossil Bees? *Mol Biol Evol* 14(10):1075-1077.
- Hedges, S. B. and Schweitzer, M. H. (1995). Detecting dinosaur DNA. *Technical Comments. Science* 268:1191.
- Austin, J. J., Ross, A. J., Smith, A. B., Fortey, R. A. and Thomas, R. H. (1997). Problems of reproducibility – does geologically ancient DNA survive in amber-preserved insects? *Proc R Soc Lond B* 264:467–474.
- Hänni, C., Laudet, V., Coll, J. and Stehelin, D. (1993). An unusual mitochondrial DNA sequence variant from an Egyptian mummy. *Genomics* 22(2):487-9.
- Caramelli, D., Lalueza-Fox, C., Vernesi, C., ... Bertorelle, G. (2003). Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc Natl Acad Sci USA*. 100(11):6593-7.
- Béraud-Colomb, E., Roubin, R., Martin, J., Maroc, N., Gardeisen, A., Trabuchet, G. and Goossens, M. Human beta-globin gene polymorphisms characterized in DNA extracted from ancient bones 12,000 years old. *Am J Hum Genet* 57(6):1267-74.
- Shapiro, B. and Hofreiter, M. (2014). A Paleogenomic Perspective on Evolution and Gene Function: New Insights from Ancient DNA. *Science* 343(6169):1236573.
- Rieder, M. J., Taylor, S. L., Tobe, V. O. and Nickerson, D. A. (1998). Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Research* 26:4.
- Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., Kivisild, T., Torroni, A. and Villems, R. (2012). A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. *Am J Hum Genet* 90(4):675–684.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*;74(12):5463-7.

Kasutatud veebiaadressid

- [1] http://www.phylotree.org/resources/rCRS_annotated.htm (viimati kasutatud 17.05.2018)
- [2] <http://www.phylotree.org/Y/tree/index.htm> (viimati kasutatud 18.02.2018)
- [3] <https://www.nature.com/scitable/topicpage/genetic-recombination-514> (viimati kasutatud 01.03.2018)
- [4] www.ancestraljourneys.org/ (viimati kasutatud 20.02.2018)
- [5] www.umap.openstreetmap.fr/ (viimati kasutatud 20.02.2018)
- [6] www.khanacademy.org/science/biology/biotech-dna-technology/dna-cloning-tutorial/a/overview-dna-cloning/ (viimati kasutatud 30.04.2018)
- [7] <https://www.neb.com/applications/library-preparation-for-next-generation-sequencing/dna-library-preparation> (viimati kasutatud 20.04.2018)
- [8] <https://www.youtube.com/watch?v=fCd6B5HRaZ8> (viimati kasutatud 03.03.2018)
- [9] <http://www.nhm.ac.uk/content/dam/nhmwww/our-science/dpts-facilities-staff/Coreresearchlabs/nanodrop.pdf> (viimati kasutatud 21.05.2018)
- [10] <https://technelysium.com.au/wp/chromaspro/> (viimati kasutatud 20.03.2018)

Lisad

Lisa 1. Sekveneeritud alad, nukleotiidsed asendused ja nende põhjal määratud mtDNA haplogrupid 12 arheoloogide mtDNA jaoks. Nukleotiidsed asendused on märgitud nii, et referentsgenoomi (Behar jt., 2012) järgi defineeritud nukleotiidsed positsiooni (np) tähistavale numbrile eelneb referentsgenoomis ja järgneb uuritud proovis esinev nukleotiid. Näiteks, G73A tähendab guaniini asendumist adeniiniga positsioonis 73. Transitsioonid on tähistatud kahe suure tähega positsiooni numbriga nende vahel (nt. T7028C), transversioonid – suure ja väikese tähega (nt. A2871t). Tühi kast tähendab, et vastavat ala antud proovi jaoks ei sekveneeritud.

Proov	Sekveneeritud ala (np)	Asendused kontrollregioonis (np 16024-576)	Asendused kodeerivas regioonis (np 577-16023)	Haplo-grupp
1	0-16569	G73A, C146T, C152T, C195T, A247G, A285T, A16129G, T16223C, G16230A, T16278C, C16311T	A769G, A825T, A1018G, G2706A, A2758G, C2885T, T3594C, G4104A, T4312C, A4793G, T7028C, G7146A, T7256C, G7379A, A7521G, T8468C, T8655C, G8701A, C9540T, G10398A, T10664C, A10688G, C10810T, C10873T, C10915T, A11719G, T12705C, G13105A, G13276A, T13506C, T13650C, T14766C	H7
2	0-720; 7857-9380; 16014-16496	C146T, C152T, C195T, A247G, A16129G, A16162G, T16223C, G16230A, T16278C, C16311T	A8271t, T8468C, T8655C, G8701A	H1a2
3	0-16569	C146T, C152T, C195T, A247G, A16129G, A16162G, T16223C, G16230A, T16278C, C16311T	A769G, A825T, A1018G, G2706A, A2758G, C2885T, G3010A, T3594C, G4104A, T4312C, T7028C, G7146A, T7256C, A7521G, A8271t, T8468C, T8655C, G8701A, C9540T, G10398A, T10664C, A10688G, C10810T, C10873T, C10915T, A11719G, A11914G, T12705C, G13105A, G13276A, T13506C, T13650C, T14766C	H1a2
4	0-718; 7833-9407; 16017-16491	C146T, C152T, C195T, A247G, A16129G, A16162G, T16223C, G16230A, T16278C, C16311T	A8271t, T8468C, T8655C, G8701A	H1a2
5	0-720; 16017-16493	C146T, C152T, C195T, C242T, A247G, C295T, C462T, T489C, C16069T, T16126C, A16129G, G16145A, T16172C, G16230A, C16260T, C16261T, T16278C, C16311T		J1b1a1
6	0-16569	C146T, C152T, C196T, A200G, A247G, T310C, G499A, A16129G,	A769G, A825T, A1018G, A1811G, A2758G, C2885T, T3594C, G4104A, T4312C, T4646C, T5999C, A6047G, G7146A, T7256C, A7521G, T8468C, T8655C, G8701A, C8818T, C9540T, G10398A, T10463C, T10664C, A10688G, C10810T, C10873T,	U4a2a2

		T16223C, G16230A, T16278C, C16311T	C10915T, C11332T, A11467G, A11914G, A12308G, G12372A, T12705C, G13105A, G13276A, T13506C, T13650C, G13759A, C14620T, T15693C	
7	0-680; 14886-15937; 16014-16493	T72C, G73A, C146T, C152T, A247G, A302C, T16223C, G16230A, T16278C, T16298C	T15454C	HV0e
8	0-16569	C146T, C152T, A189G, C195T, A247G, T16126C, A16129G, T16223C, G16230A, T16278C, C16294T, C16269T, T16304C, C16311T	G709A, A769G, A825T, G930A, A1018G, G1888A, A2758G, C2885T, T3594C, G4104A, T4216C, T4312C, A4917G, G5147A, G7146A, T7256C, A7521G, T8468C, T8655C, G8697A, G8701A, C9540T, G10398A, T10463C, T10664C, A10688G, C10810T, C10873T, C10915T, A11251G, A11812G, A11914G, T12705C, G13105A, G13276A, G13368A, T13506C, T13650C, A14233G, G14905A, C15452a, A15607G, G15928A	T2b
9	0-719; 7173-8357; 16012-16493	C146T, C152T, A189G, C198T, T204C, G207A, A247G, A385G, A16129G, G16230A, T16278C, C16292T, C16311T, G16346A	G8251A, C7864T, T7521C, A7256G	W1
10	0-698; 16014-16491	C146T, C152T, A189C, C195T, A247G, T16126C, A16129G, T16223C, G16230A, T16278C, C16294T, C16296T, T16304C, C16311T		T2b
11	0-726; 6468-7319; 12540-12778; 16017-16493	G73A, C146T, C152T, C195T, A247G, C16083T, A16129G, G16230A, T16278C, C16311T	T7028C, G7146A, T7256C, C12705T	H
12	0-726; 6468-7319; 12540-12778; 16017-16493	G73A, C146T, C152T, C195T, A247G, C16083T, A16129G, G16230A, T16278C, C16311T	T7028C, G7146A, T7256C, C12705T	H

Lisa 2. Sekveneeritud alad, nukleotiidsed asendused ja nende põhjal määratud mtDNA haplogrupid kuue täielikult sekveneeritud vana mtDNA jaoks, mille võrdlusanalüüsil arheoloogide mtDNAGA selgus, et need langevad samasse või väga lähedasse haplogruppi. Nukleotiidsed asendused on märgitud nii, et referentsgenoomi (Behar jt., 2012) järgi defineeritud nukleotiidsed positsiooni (np) tähistavale numbrile eelneb referentsgenoomis ja järgneb uuritud proovis esinev nukleotiid. Näiteks, G73A tähendab guaniini asendumist adeniiniga positsioonis 73. Transitsioonid on tähistatud kahe suure tähega positsiooni numbriga nende vahel (nt. T7028C), transversioonid – suure ja väikese tähega (nt. A2871t), insertioonid – punktiga pärast inserteerumise positsiooni (nt. T302.1C).

Proov	Asendused kontrollregioonis (np 16024-576)	Asendused kodeerivas regioonis (np 577-16023)	Haplo-grupp
1	A73G, A263G, T16126C, C16294T, C16296T, T16304C	G709A, A750G, G930A, A1438G, C1888A, A2706G, G4491A, A4917G, G7565A, G8020A, G8697A, T9861C, T10463C, A11251G, G11719A, C11735T, A11812G, G13368A, A14233G, C14766T, G14905A, A15326G, G15452A, A15607G, G15928A	T2b
2	A73G, A263G, T16126C, C16294T, C16296T, T16304C, T16519C	G709A, A750G, G930A, A1438G, A2706G, G4127A, A4917G, G8697A, T10463C, A11251G, G11719A, A11812G, G13368A, C14233G, C14766T, G14905A, C15326G, G15452A, A15607G, G15928A	T2b
3	A73G, A263G, T302.1C, T310.1C, T16126C, C16294T, C16296T, T16304C, T16519C	G709A, A750G, G930A, A1438G, G1888A, A2706G, A4917G, C7610T, G8697A, T9077C, T10463C, A11251G, T11719A, A11812G, A12030G, G13368A, A14233G, C14766T, G14905A, A15326G, G15452A, A15607G, G15928A	T2b
4	A73G, A263G, T310.1C, T16126C, C16294T, C16296T, T16304C, T16519C	G709A, A750G, G930A, A1438G, G1888A, A2706G, A4917G, G8697A, G9663A, T10463C, A11251G, G11719A, A11812G, G13368A, A14233G, C14766T, G14905A, A15326G, G15452A, A15607G, G15928A	T2b
5	A73G, A189G, T195C, A263G, G499A, C16223T, T16356C, T16519C	A750G, A1438G, A1811G, A2706G, T5999C, A6047G, G7566A, A9803G, C11332T, A11467G, G11719A, A12308G, G12372A, C14620T, C14766T, A15326G, T15693C	U4
6	A73G, A263G, C295T, C462T, C16069T, T16126C, C16261T	A750G, A2706G, G5460A, A10398G, T10667C, A11251G, A12612G, G12795A, G13708A, T13879C, C15326G, A15442G, G15452A	J1b1a3

LIHTLITSENTS

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Maria Maloverjan
(sünnikuupäev 14.09.1996),

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

“Vanas DNAs leiduva saastuse tuvastamine vanade ja kaasaegsete mitogeenide võrdlusanalüüsi abil”,

mille juhendajad on Alena Kushniarevich ja Kristiina Tambets,

1.1 reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2 üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu alates **22. 05. 2018** kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 22.05.2018